Development of synthetic biology tools and construction of Yarrowia lipolytica chassis strains for aromatic molecules production

Macarena Larroude

HAL Id: tel-03243034
https://pastel.hal.science/tel-03243034

Submitted on 31 May 2021

Development of synthetic biology tools and construction of *Yarrowia lipolytica* chassis strains for aromatic molecules production

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°581 : agriculture, alimentation, biologie, environnement et santé (ABIES)
Spécialité de doctorat : Biotechnologies
Unité de recherche : Université Paris-Saclay, INRAE, AgroParisTech, Micalis Institute, 78350, Jouy-en-Josas, France
Référent : AgroParisTech

Thèse présentée et soutenue à Paris, le 20 mai 2020, par

Macarena LARROUDÉ

Composition du Jury

Matthieu JULES Professeur, AgroParisTech	Président
Paola BRANDUARDI Professeur, Università di Milano-Bicocca	Rapporteur & Examinatrice
Patrick FICKERS Professeur, Université de Liège	Rapporteur & Examinateur
Florence BORDES Maître de Conférence, INSA Toulouse	Examinatrice
Ioana POPESCU Maître de Conférence, Université d'Évry-Val-d'Essonne	Examinatrice
Tristan ROSSIGNOL Chargé de recherche (HDR), INRAE (MICALIS)	Directeur de thèse

Acknowledgements

It is surprising how fast these last three years have passed, and how amazing they have been. Thus, I would like to thank the people that accompanied me during this time.

I would like to express my deepest thanks and gratitude to my PhD advisor, Tristan Rossignol, for his mentoring, support and constructive feedbacks during this time. Also, for trusting me and giving me the freedom to develop my work. Thank you very, very much! I would also like to thank very much Jean-Marc Nicaud, director of BIMLip team, who was always available for discussions, ideas and shearing his knowledge.

I want to thank Paola Branduardi and Patrick Fickers for accepting being reporters of this work, and also Florence Bordes, Ioana Popescu et Matthieu Jules for accepting being part of the jury. Thank you for your time and the feedbacks you gave me of my work. I really appreciated the discussions we had during the defence of this work. Also, thanks to Verena Siewers and Mislav Oreb for their advices as external advisors of my PhD project, it was always nice and useful to talk to you; and all ABIES team, it is for sure the best doctoral school we can wish to be part of.

I had the chance to do my work in the frame of a large consortium project. It was a great experience and I am endlessly grateful to all CHASSY members. Being part of the European CHASSY project gave me the opportunity of meeting awesome people, great scientist and growing professionally and personally in a really motivating environment.

Certainly, the work was very pleasant thanks to all the people I had the chance to work with daily. I thank all BIMLip members I had the chance to meet since I arrived at the team. I would like to specially thank Pauline, with whom I've shared several hours of work and fun, Young, for her kindness and caring, Léa for being always in a good mood and ready to help. Big thanks also to Rodrigo and Heber, who were always there helping and giving good advices while I was doing my first steps into Yarrowia's world. And, of course, thanks to all people that are part of the Bâtiment 526 at INRAE-Jouy. Thanks for your encouraging words, your kindness, the coffee breaks, the discussions, the carpooling when trains were cancelled. Thanks to Vincent who taught me all about HPLC and drove me during the SNCF strike. And many thanks to Abarna, who is always

smiling, in good mood, caring about others. Thanks for your great company on Vauboyen road and RERC, it is great to have you around :)

Finally, I would like to thank my family and friends, most of which are at more than 10.000 Km away and I cannot see as much as I would like, but they are always present and available for messages, calls and videocalls.

Thanks to Angélica, Tatiana, Fanny, Tomás, Bianca, Guille and Youri, for being part of my daily life. For the company, the talks, the walks, the food… for being always there.

To the ones that are very far away, thank you for being present no matter how. Thanks to Aye, Lu, Ana and Sol, that are always calling and asking how and what am I doing, thanks for the caring and the presence. To Na, Flaqui, Otis and Glen, because I now they are there. My dearest Chilita, so great to have you in my life since ever. I'm extremely grateful that life continues to put us close even without planning, and in a so unexpected place and time. Thanks for the talks, for being, for sharing.

A mis tan adoradas abuelas, a mis tio(a)s y primo(a)s y a los Figueroa. Gracias infinitas por todo el amor desmesurado que recibo de ustedes desde siempre! No puedo explicarles lo linda que hacen mi vida!

To Fernando, who walks with me every day full of energy and smiles. I cannot thank you enough for all your support and love. A toda la flia. Brufal, gracias por todo el cariño y apoyo.

To Ma, Pa, Mar y Bau, for the love and unconditional and infinite support. Gracias infinitas por ser mi piedra angular y por estar siempre tan cerca. Los adoro!

Table of contents

x

x

List of figures.

List of tables.

List of abbreviations.

- **2PE**, 2-phenylethanol
- **4OHPAA**, 2-(4-hydroxyphenyl)acetate
- **4OH2PE**, 2-(4-hydroxyphenyl)ethanol
- **AAA**, Aromatic amino acid
- **bp**, base pair(s)
- **CDW,** cell dry weight
- **CRISPR,** clustered regularly interspaced short palindromic repeats
- **DSB,** double strand break
- **DBTL,** design, build, test and learn
- **E4P**, erythrose-4-phosphate
- **GG**, Golden Gate
- **GGA,** Golden Gate Assembly
- **GRAS**, generally recognized as safe
- **gRNA,** guide RNA
- **HR**, homologous recombination
- **IPA**, indole-3-pyruvic acid
- **kb**, kilobase(s) or 1000 bp
- **NCR**, nitrogen catabolite repression
- **NHEJ**, non-homologous end joining
- **nt**, nucleotide(s)
- **OD$_{600}$**, optical density at 600 nm
- **ORF**, open reading frame
- **PAA**, 2-phenylacetate
- **PAM,** protospacer adjacent motif
- **PCR**, polymerase chain reaction
- **PEP**, phosphoenolpyruvate
- **Phe**, L-Phe, phenylalanine
- **PPP,** pentose phosphate pathway
- **PVA**, protodeoxyviolaceinic acid

- **rDNA**, ribosomal RNA gene
- **TCA,** tricarboxylic acid
- **TF,** transcription factor
- **Trp**, L-Trp, tryptophan
- **TU,** transcription unit
- **Tyr**, L-Tyr, tyrosine
- **UAS,** Upstream activating sequences
- **YNB**, minimal Yeast Nitrogen Based media
- **YPD**, rich media Yeast extract Peptone Dextrose

1 Introduction

1.1 Context of the study

1.1.1 Synthetic biology, metabolic engineering and chassis strains

Nowadays, valuable chemicals and materials are mostly manufactured from petroleum-derived sources by chemical synthesis. However, these production methods usually require toxic solvent, harsh conditions, and generate by-product wastes, which together with the limited availability of fossil resources, render the processes unsustainable. Therefore, the development of economically feasible and sustainable processes as alternatives to oil-based chemistry is a major industrial goal.

In this regard, the development of microorganisms capable of producing valuable molecules at industrial levels from renewable resources provides an attractive alternative. This objective can be address today by the combined endeavour of metabolic engineering and synthetic biology disciplines.

Even though highly close disciplines and concepts, metabolic engineering focus on engineering cells as factories to overproduce both native and non-native metabolites by the rewiring of basic metabolism, while synthetic biology aims to design and construct new biological systems by providing parts or building blocks that are standardized, well-characterized and which can be assembled to construct a well-defined metabolic pathway or even a cell with an entire synthetic genetic material (Adams, 2016; Nielsen et al., 2014; Stephanopoulos, 2012).

The core of synthetic biology is synthetic DNA and its use applying engineering strategies to biology. Thus, it uses standardized building blocks to rationally design, synthesize, and build genetic networks in a modular fashion for customized biological functions. To that end, several synthetic biology tools such as Biobricks (Knight, 2003), Golden Gate (Engler et al., 2008), Gibson Assembly (Gibson et al., 2009) and CRISPR-Cas9 (Jinek et al., 2012) have been developed and proposed to achieve metabolic engineering plans in a much faster, efficient and cost-effective way.

Nevertheless, the modification of endogenous metabolism or the addition of heterologous pathways interfere with the complex metabolic network and regulatory mechanisms of the cell. This may result in changes in carbon source utilization, energy transfer changes and may affect cell growth (Lyu et al., 2018). Despite recent progress, our knowledge of how complex metabolic interactions are regulated is still limited and challenges the development of cell factories. Therefore, in order to achieve the construction of systems that meet the expected requirements, e.g. a production yield or strain stability, an iterative synthetic biology workflow was proposed: the DBTL cycle (Design-Build-Test-Learn cycle) (Nielsen and Keasling, 2016) (Figure 1). This cycle intends to iteratively improve a biological process, such as the enhancement of a producer strain, the expansion of the catalytic activity or the stability of a protein. It usually begins with a *design* step, where the desired strategy is decided based on *in silico* analysis using modelling and databases. Afterward, during the *build* step, the DNA constructions determined from the design step are built and used to engineer the organisms. The obtained new biological systems are then tested. The *test* step compares if the output results fit with those expected from the design step. Finally, the *learn* step focus on obtaining leverage information from the new biological data, such as regulation patterns that can help to improve current models and the following design round. Thus, engineering a cell factory involves several rounds of the DBTL cycle. A review describing each part of the DBTL cycle with examples of applications to the production of esterases and lipases in *Y. lipolytica* (Gamboa-Melendez et al., 2018) is presented in the Annexes section (6.2).

In addition to the development of technologies and tools, another important point to consider when working on microbial-based biotechnology, is the choice of the organism to work with, the chassis. In classical engineering terms, a chassis is the framework that supports other physical components needed for the construction and function of the structure. In synthetic biology, a chassis refers to the organism that houses genetic components and supports them by providing the resources to function, such as transcription and translation machinery (Adams, 2016; Chi et al., 2019). Synthetic biology was mostly developed on domesticated, well-characterized laboratory strains such as *Escherichia coli* or *Saccharomyces cerevisiae*. However, for expanding the current scope of synthetic biology it is critical to develop a panel of chassis organisms that can

thrive under a variety of conditions. Several key considerations must be taken into account when choosing a strain to work with. In this sense, the basic knowledge of the laboratory strains is helpful to facilitate the development of chassis and also to design accurate *in silico* models; the availability of optimized transcription and translational control elements should be considered in order to tightly control gene expression; and the readiness of tools needed to introduce the designed genetic circuits into chassis will ease the construction and save time in the process. A good strain selection can, indeed, facilitate the development of a project (Adams, 2016; Brophy and Voigt, 2014).
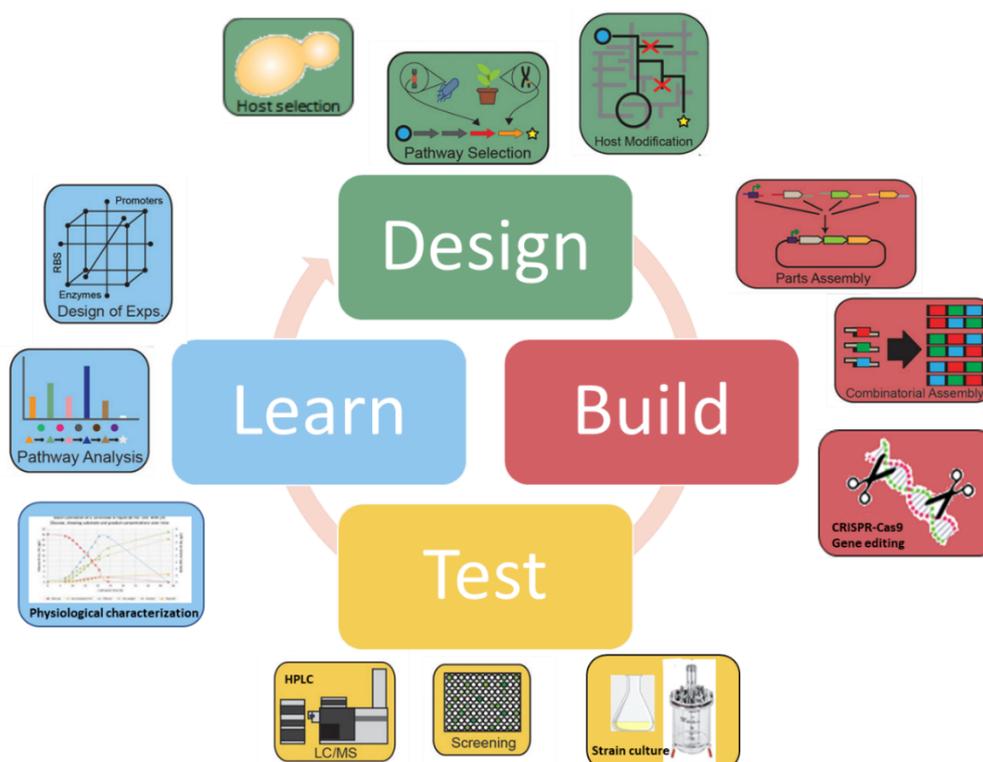
Adapted from Petzold et al., 2015

Figure 1. Design-Build-Test-Learn cycle for metabolic engineering, highlighting important parts of each component.

The Design component identifies the problem, selects the desired pathway and host; the Build component selects, synthesizes, and assembles parts for incorporation into the host; the Test component validates the engineered strains for target molecule production, transcripts, proteins, and metabolites; the Learn component analyses the Test data and informs subsequent iterations of the cycle. Figure adapted from Petzold et al., 2015

Furthermore, when constructing cell factories, a large phenotypic space needs to be explored in order to identify high producers. Thus, the construction of libraries of strains harbouring different combination of DNA parts (e.g. genes and promoters) is an efficient way to quickly have a large set of phenotypes to screen. In consequence, the availability of appropriate high-throughput screening techniques is of great value. In this sense, highly sensitive biosensors are helpful for screening the resulting libraries. Most of these systems are based on transcription factors (TFs) that detect a molecule of

interest and, in response, activate the expression of a reporter gene that produces an easily detectable signal, e.g. fluorescence. However, most biosensors reported to date are based on a small number of well-characterized TFs, thus, limited to a restricted number of detectable molecules. Besides, these systems are developed in conventional strains such as *E. coli* and *S. cerevisiae* (Cheng et al., 2018; Mahr and Frunzke, 2016). The identification of new sensor components, the extension or the transfer to other organisms as biosensors are key components to setup robust circuits for synthetic biology.

Synthetic biology and metabolic engineering have advanced quickly during the last decades thanks to the reduction of cost of sequencing and DNA synthesis and the merging of large dataset with modelling and novel bioinformatics tools. Additionally, academic competitions such as iGEM (the international Genetically Engineered Machine competition) have helped to raise awareness and attract young pioneers to the field. As a result, great advances were done in a large scope of domains, from production of chemicals and treatment of polluted environments (Khalil and Collins, 2010) to medical applications (Weber and Fussenegger, 2012), including reprogramming immunity cells to target cancer cells (Wu et al., 2019). These microbial-based technologies, generally called 'white biotechnologies', can provide an environmentally-friendly process, with less energy consumed and less wastes generated, to produce molecules of interest by using low-cost, non-food resources (Heux et al., 2015). This is reflected in a global market valued at US$ 203.28 billion in 2015, and expected to reach approximately US$ 727.1 billion by 2024, according to a report by Grand View Research, Inc. (Grand View Research, 2016a). Nevertheless, and despite efforts, there is still a long way to go to have complete biotechnological production of the nowadays petroleum-derived compounds. Therefore, the development of rapid engineering of readily controllable synthetic hosts is a need.

1.1.2 The CHASSY project

The work done in this PhD thesis and presented in this manuscript is part of the European CHASSY project (https://chassy.eu), a collaboration between academics and industrial partners, financed by the European Union's Horizon 2020 research and innovation program under grant agreement No 720824.

The aim of CHASSY is the development of versatile, efficient and robust chassis strains for it use as cell factories. This objective will be addressed by building yeast

platforms with optimized levels of core precursors needed for the production of a wide variety of high-value oleochemicals and aromatic molecules. The knowledge gained from systems biology together with the developed engineering tools of synthetic biology is used to redesign metabolic pathways in the yeasts *S. cerevisiae*, *Yarrowia lipolytica* and *Kluyveromyces marxianus*.

1.1.3 **Objectives of the thesis**

The construction of platform biological systems with an optimized supply of precursor to be used as building blocks for the synthesis of a wide range of high value products will make bio-based production by microorganisms more time and cost efficient. Additionally, the use of non-conventional organism, such as the yeast *Y. lipolytica*, as cell factories can be beneficial due to particular traits compared to the traditional *E. coli* and *S. cerevisiae*. However, more efficient tools than the existing ones, enabling the rapid construction of engineered strains need to be developed for these non-conventional organisms.

The aim of this thesis is the construction of tailor-made chassis strains optimized for synthesis of aromatic amino acids (AAA), which are key precursor molecules for commercially valuable compounds, in the industrial non-conventional yeast *Y. lipolytica*. To accelerate the construction of strains, synthetic biology tools dedicated to this yeast are aimed to be developed as well.

This will be achieved with the development of synthetic biology tools that can be applied in a generic systematic manner in *Y. lipolytica*. More precisely, a technique to assemble DNA parts in a fast and efficient fashion and a technique to increase the capacity to modify the genome of this yeast will be designed, constructed and set up. Thereafter, the newly developed molecular tools will be used for the construction of rationally engineered strains optimized for the synthesis of key precursor molecules, AAAs. The production of AAAs will be compared among the constructed strains in order to identify the best producer, which will be considered as the chassis strain. Finally, the potential of the chassis strain will be evaluated by the *de novo* biosynthesis of several value-added compounds derived from the phenylalanine (Phe) and tyrosine (Tyr). To this end, further engineering of the chassis strain will be necessary to express heterologous pathways and/or regulate expression of endogenous genes.

The AAA pathway has been poorly explored in *Y. lipolytica* up to now. Therefore, the optimization of this pathway, in addition to enlarge the scope of possible applications, will increase the knowledge on how this metabolic pathway may function in this yeast. The acquired informations can be then used to feed the DBTL cycle and improve next generation of aromatics cell factories and experiments design.

1.2   *Yarrowia lipolytica* **as a chassis strain**

As mentioned before, several key considerations have to be taken into account when choosing a strain to work with. The basic knowledge, available techniques and the physiological characteristics are important criteria to look at. In this regard, the generally referred as non-conventional yeast *Y. lipolytica*, appears as a good candidate to be developed as a high industrial-value chassis strain.

1.2.1   **General characteristics of *Y. lipolytica***

This strictly aerobic yeast has a broad substrate range due to its potential of producing lipases and proteases, consequently, it can be found readily in nature. It is found primarily in foods with high proportion of fat and/or protein such as cheeses, yoghurts and sausages, and it has also been isolated from various environments such as lipid-rich media (e.g. oil-polluted media) or marine and hypersaline environments (Groenewald et al., 2014). Furthermore, this yeast is unable to grow above 32°C and, therefore, considered non-pathogenic which is particularly valuable for industrial processes.

Y. lipolytica can efficiently use hydrophobic substrates (Fickers et al., 2005) and naturally accumulate more than 20% of its dry weight in lipids, thus, it is considered as an oleaginous microorganism (Nicaud, 2012). Furthermore, it was describe to have a strong endogenous flux through tricarboxylic acid (TCA) and pentose phosphate (PPP) pathways (Christen and Sauer, 2011; Wasylenko et al., 2015), being the oxidative PPP the primary source of NADPH supporting the lipogenesis in *Y. lipolytica* (Liu et al., 2016). Consequently, this yeast has naturally high acetyl-CoA and malonyl-CoA pools (Christen and Sauer, 2011).

Natural isolates of *Y. lipolytica* appear to have widely divergent structures in terms of physiology and genetics (Egermeier et al., 2017; Naumova et al., 1993). In terms of physiology, wild-type strains of *Y. lipolytica* exhibit diverse colony morphologies that can range from heavily convoluted and matt to smooth and glistening depending on

growth conditions (aeration, carbon and nitrogen sources, pH, etc.) and genetic backgrounds (Egermeier et al., 2017; Nicaud, 2012). Furthermore, this hemiascomycetous yeast is a natural dimorphic fungus which forms yeast cells, pseudohyphae and true hyphae in different proportions depending on the environmental conditions (Barth and Gaillardin, 1997; Thevenieau et al., 2009). In addition, these strains can have different metabolic patterns, for instance, the natural production of α-ketoglutarate by H222 (Yovkova et al., 2014), or high citric acid production by A101 (Wojtatowicz and Rymowicz, 1991).

Regarding genomics, the most widely used *Y. lipolytica* genetic backgrounds are the wild-type French strain W29 (CLIB89), the wild-type German strain H222 (DSM 27185), the wild-type American strain CBS6124–2, and the wild-type Polish strain A101 (Barth and Gaillardin, 1996; Nicaud, 2012; Wojtatowicz and Rymowicz, 1991). Yet, the reference genome for *Y. lipolytica* is the E150 (CLIB122) strain, derived from multiple back-crosses between the W29 and CBS6124–2 strains (Barth and Gaillardin, 1996). This yeast has a 20.5Mb genome composed of six chromosomes, it has been sequenced and deeply annotated (Dujon et al., 2004; Magnan et al., 2016). The genome sequences of several strains, including W29, are publicly available on NCBI or GRYC databases (http://gryc.inra.fr/). In addition, the sequence of the mitochondrial genome with a size of 47.9 kb was reported (Kerscher et al., 2001). Most of the natural isolates strains are haploid, and there are not natural plasmids in *Y. lipolytica* (Thevenieau et al., 2009). However, strains differ on their chromosome length polymorphism and the presence of retrotransposon Ylt1 that has only been detected in the American strain (Schmid-Berger et al., 1994; Thevenieau et al., 2009).

The retrotransposon Ylt1 is characterized by long terminal repeats (LTR) termed zeta (ζ) at each side. Zeta elements have a well conserved sequence, can exist as part of retrotransposon Yltl as well as a solo element in the genome (Schmid-Berger et al., 1994), and can be used for exogenous DNA integration into the genome (Pignède et al., 2000). As mentioned, the number of copies of Yltl and solo zeta depends on the strain, ranging from more than 30 per haploid genome to any (Barth and Gaillardin, 1997; Mauersberger et al., 2001). Integration cassette flanked by zeta integrates randomly into the chromosome in a strain devoid of zeta, while, in a strain containing zeta, insertion occurs mainly at a zeta locus (Pignède et al., 2000).

Another important feature in terms of *Y. lipolytica* genomics, with implications on the rational genome-editing of this yeast, is its DNA double strain break repair mechanism. *Y. lipolytica* preferentially uses non-homologous end joining (NHEJ), rather than homologous recombination (HR) (Richard et al., 2005). HR requires long (~1 kb) homologous flanking fragments to occur in this yeast, still, with a low frequency of proper integration. Regarding NHEJ, a principal component of this mechanism is the ku70/ku80 heterodimer. Both ku70 and ku80 bind to the broken DNA ends and have bridging activity regardless the sequence homology of the broken ends (Lustig, 1999). Hence, the disruption of *KU70* gene significantly hinders NHEJ efficiency and increases the use of HR (Kretzschmar et al., 2013; Verbeke et al., 2013). Therefore, Δ*ku70* strains are commonly used for targeted gene insertion in *Y. lipolytica*.

1.2.2   *Y. lipolytica* **as a biotechnological workhorse**

In addition to fundamental research on hydrophobic substrates utilization, peroxisome biogenesis, lipid metabolism, protein secretion mechanism and dimorphism (Barth et al., 2003; Nicaud, 2012), *Y. lipolytica* proved to have potential to act as a biotechnological workhorse.

This non-conventional yeast, is of strong interest thanks to its capacity of use cheap carbon sources, its "Crabtree-negative" condition, its high tolerance to a variety of organic compounds, salt concentrations and pH levels, and particularly to its capacity to efficiently use hydrophobic substrates and accumulate lipids (Barth and Gaillardin, 1997; Miller and Alper, 2019). Moreover, several *Y. lipolytica* derived products have been classified as Generally Regarded As Safe (GRAS) by the American Food and Drug Administration (FDA) (Groenewald et al., 2014), and very recently its biomass has also been considered as safe by the European Food Safety Authority (EFSA) (Turck et al., 2019).

Its oleaginous feature makes it especially suitable as a host organism for producing large amounts of lipids (Darvishi et al., 2017; Friedlander et al., 2016; Park et al., 2018). Thus, it has been used to produce, for instance, oleic acid oil (Tsakraklides et al., 2018); unusual fatty acids such as hydroxy fatty acids, ricinoleic acid and cyclopropane fatty acids (Beopoulos et al., 2014; Czerwiec et al., 2019); conjugated fatty acids such as conjugated linoleic acids (Imatoukene et al., 2017; Zhang et al., 2013, 2012); and

omega-6 and omega-3 fatty acids like EPA (eicosapentaenoic acid, C20:5) (Xie et al., 2015; Xue et al., 2013).

Y. lipolytica also gained popularity as an industrial host for the production of re-combinants proteins, such as lipases (Pignède et al., 2000) and α-amilases (Celińska et al., 2018, 2015a); and other chemicals, including, methyl ketones (Hanko et al., 2018), polyhydroxyalkanoates (PHAs) (Gao et al., 2015; Haddouche et al., 2011, 2010; Li et al., 2017; Rigouin et al., 2019), carotenoids (S. Gao et al., 2017; Larroude et al., 2017), erythritol (Carly et al., 2017b), mannitol (Rakicka et al., 2016), organic acids (Blazeck et al., 2015; Markham and Alper, 2018), and aroma molecules like 2-phenylethanol (rose-like odour) (Celińska et al., 2013) and γ-decalactone (peach-like odour) (Braga and Belo, 2016).

Moreover, *Y. lipolytica* can grow on cheap substrates. It can naturally grow on glycerol, and after engineering, on sucrose, starch, inulin, cellobiose, or other waste products available for reuse (Cui et al., 2011; Johnravindar et al., 2018; Lazar et al., 2013; Ledesma-Amaro and Nicaud, 2016a; Magdouli et al., 2017; Mirończuk et al., 2016; Spagnuolo et al., 2018). As well, strategies to facilitate product extraction (Ledesma-Amaro and Nicaud, 2016a) were developed, which can be of great importance to decrease bioprocess cost. This non-conventional oleaginous yeast is also a valuable candidate for environmental applications, as it was shown to be applicable for bioremediation of contaminated environments (Bankar et al., 2009).

Due to its industrial potential, several strains were constructed as platforms for specific uses. For instance, a strain with deleted extracellular proteases (AEP and AXP) for heterologous protein expression (Madzak et al., 2000), a strain in which *KU70/KU80* was deleted (*ΔKu70-ΔKu80*) to increase homologous recombination (HR) efficiency (Verbeke et al., 2013), the incorporation of a zeta docking platform to facilitate the incorporation of zeta-based integrative vectors (Bordes et al., 2007), and the deletion of the three main lipase-encoding genes (Lip2, Lip7 and Lip8) generated a strain suitable for engineer lipid metabolic pathway (Bordes et al., 2007). The selection of these engineered and application-specific derivative strains can be of interest when choosing a strain for a specific project.

The wide range of applications of *Y. lipolytica* were well and extensively reviewed recently (Darvishi et al., 2018; Ledesma-Amaro and Nicaud, 2016b; H.-H. Liu et al., 2015; Madzak, 2015).

The availability of *Y. lipolytica* complete genome sequence and, more recently, new synthetic biology tools, has allowed more extensive engineering of this yeast species, which combined with its specific physiological, metabolic and genomic characteristics, are making this yeast an important industrial cell factory host for the production of enzymes, oils, fragrances, surfactants, cosmetics and pharmaceuticals.

1.3   **Synthetic biology tools for *Y. lipolytica***

Synthetic biology is an emerging discipline that aims to apply engineering principles to biological systems to render them more controllable, standardized, and predictable. In order to accomplish these tasks and significantly accelerate research, new and highly efficient tools are being developed, allowing systems of interest to be constructed more rapidly and at a lower cost. In recent years, different synthetic biology tools have been created and applied in *Y. lipolytica*, including 1) DNA parts for the construction of expression cassettes, 2) DNA assembly techniques, 3) genome-editing techniques, and 4) computational tools. The significantly expansion of the dedicated genetic toolbox for this oleaginous yeast further expand its range of applications.

In terms of DNA parts for **expression cassettes**, promoters, terminators, selection markers and sequences for genome integration need to be considered. Selecting promoters according to their strength is the most widespread method for controlling gene expression in metabolic engineering techniques, which is a key parameter to optimize in pathway engineering. Consequently, significant efforts have been made to develop promoters exhibiting a wide range of transcriptional activities. In *Y. lipolytica*, in addition to endogenous constitutive and inducible promoters, hybrid promoters have been developed mostly by fusing multiples UAS (Upstream Activating Sequence) to a core promoter (Madzak et al., 2000; Müller et al., 1998; Park et al., 2019; Trassaert et al., 2017). These hybrid promoters, show an increase in promoter strength as a function of the number of tandem UAS; some can increase expression efficiency eight-fold, compared to the known endogenous promoter in *Y. lipolytica* (Blazeck et al., 2011). On the other hand, terminators have been much less studied and developed, even though, they are essential for completing the transcription process. Native terminators are mostly

used, however, synthetic terminators were recently constructed (Curran et al., 2015; Wagner and Alper, 2016).

Regarding selection markers, both auxotrophy and dominant markers are available for *Y. lipolytica*. Auxotrophy markers, which can only be used with specific strains, remain the best choice for performing selection in *Y. lipolytica* (e.g., leucine, uracil, lysine, or adenine) (Barth and Gaillardin, 1996). On the other hand, dominant markers available for *Y. lipolytica* include the hygromycin resistance gene from *E. coli* (Cordero Otero and Gaillardin, 1996), the nourseothricin resistance gene from *Streptomyces noursei* (Kretzschmar et al., 2013), the mycophenolic acid resistance gene from *E. coli* (Wagner et al., 2018), and the zeocin resistance gene from *Streptoalloteichus hindustanus* (Tsakraklides et al., 2018). Other dominant markers involve the utilization of a specific carbon source, such as sucrose when the *SUC2* gene from *S. cerevisiae* is expressed (Nicaud et al., 1989), erythritol when the *EYK1* gene is expressed (Vandermies et al., 2017) or acetamide by expressing the gene *AMD1* gene (Hamilton et al., 2020). For the last two cases, deletion of these genes is required in order to be used afterwards as markers, as for auxotrophy markers.

Finally, integration into the genome is mostly achieved by large (0.5–1 kb) homologous 5′ and 3′ flanking regions or by zeta regions from the Ylt retrotransposon. While homologous sequences are used for specific genome targets, zeta sequences are involved in random integration into the genome in strains devoid of retrotransposon elements (Juretzek et al., 2001). For the purposes of heterologous expression and genetic engineering, the integration into the genome is preferred rather than synthetic replicative plasmids (Nicaud et al., 2002). The use of such vectors is limited in *Y. lipolytica* because of low copy numbers (~1–3 plasmids/cell) and the high frequency of loss (Fournier et al., 1993; Nicaud et al., 1991).

As the cornerstone of synthetic biology, the **DNA assembly** process allows the construction of novel biological systems and devices using defined, standardized, and well-characterized components. The assemblies consist of DNA fragments physically linked end to end, creating a target higher-order assembly that is then joined to a vector. Traditional techniques employing restriction digestion and element-by-element cloning are time consuming and cost inefficient (Celińska and Grajek, 2013; Matthäus et al., 2014). Therefore, significant efforts are being made to develop methods that would

allow multigene cassettes to be constructed more quickly and efficiently, which will consequently enable the easier construction of strains with complex genetic functionalities. Among the most recent and relevant ones are Golden Gate assembly, Gibson assembly, BioBricks and Gateway cloning (Figure 2).

The Golden Gate (GG) is a modular cloning approach, based on type IIS restriction enzymes, for assembling multiple genes via a single-step, one-pot reaction (Engler et al., 2008). Type IIS enzymes cut outside their recognition sites to excise sequences with overhangs that can be, thus, arbitrarily defined. The DNA modules are designed in such a way that the enzyme recognition site is lost after the digestion and the overhangs generated are compatible to each other to allow a predetermined sequential order of the modules. In *Y. lipolytica*, this technique has been successfully used to assemble heterologous pathways of three genes at once, allowing the production of β-carotene and the utilization of xylose as carbon source (Celińska et al., 2017; Larroude et al., 2019). One of the advantages of GG is that it allows combinatorial assembly, which can be used to efficiently generate libraries, and was used to identify the best promoter-gene pairs for the carotenoids pathway (Larroude et al., 2017).

Gibson assembly allows multiple DNA fragments to be assembled regardless of their length or end compatibility. It exploits three different enzymes: (1) The exonuclease that creates single-stranded 3′ overhangs that facilitate the annealing of fragments that share complementarity at one end (overlap region), (2) the DNA polymerase that fills gaps within each annealed fragment, and (3) the DNA ligase that seals nicks in the assembled DNA. The end result is a double-stranded fully sealed DNA molecule (Gibson et al., 2009). In *Y. lipolytica*, this technique was used to study the xylose pathway (Rodriguez et al., 2016).

A set of BioBrick-based vectors dedicated for *Y. lipolytica* was developed by Wong and colleagues (Wong et al., 2017). The so called YaliBricks comprises four compatible restriction enzyme sites (*Avr*II, *Xba*I, *Spe*I, and *Nhe*I) that enable modular genetic engineering and the reuse of parts. The system was used for constructing the violacein biosynthetic pathway (Wong et al., 2017). It is a fast and easy method, but it relies on specific restriction-site-free genes.

The Entry/Gateway® method employs site-specific recombination between att sites on interacting molecules to rapidly clone single DNA sequences in multiple destination plasmids (Hartley et al., 2000). It was recently adapted for use in *Y. lipolytica* and used for the construction of a library overexpressing individual transcription factors to identify regulators involved in lipid metabolism (Leplat et al., 2018).

It is important to note that there are no perfect techniques for universal DNA assembly. Each method presents advantages and disadvantages in different situations. Therefore, the selection of a method will depend on one's objectives.
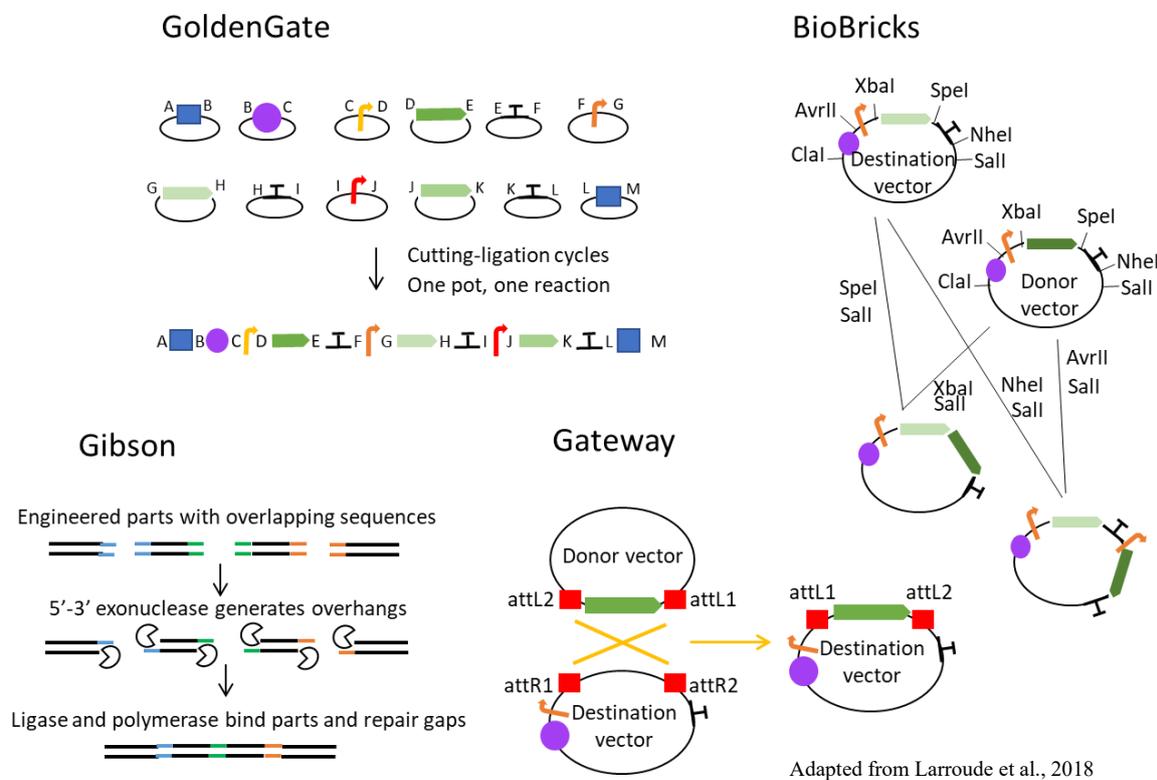
Figure 2. DNA assembly techniques for *Y. lipolytica*.

Golden Gate assembly exploits type IIS restriction enzymes, which cut outside their recognition sites to excise parts with arbitrary defined overhangs. Through the careful selection of compatible overhangs, such parts can be assembled altogether in a defined order. In the figure, the letters A to M represent different compatible 4-nt overhangs; the yellow, orange, and red arrows represent promoters; the green arrows represent genes; the violet circles represents markers; the blue squares represents insertion sequences; and the Ts represents terminators. The **BioBricks** technique is used to clone parts via restriction digestion and subsequent ligation of the resultant compatible sticky ends. YaliBricks vectors were designed to have AvrII, XbaI, SpeI and NheI endonuclease site recognition. The ligation of the compatible overhangs produces a scar that is no longer recognized by either enzyme. In **Gibson** assembly, parts are synthetized to overlap by 30+ bp. Their ends are then processed by an exonuclease that creates single-stranded 3' overhangs, which facilitates annealing. The overhangs are fused together using a polymerase, which fills in gaps within each annealed fragment, and a ligase seals gaps in the assembled DNA. In the **Gateway** method, the gene of interest, which has been cloned into the entry vector, is transferred into the destination vector via att site recombination. The expression vector obtained is digested to release the expression cassette and used to transform *Y. lipolytica*. Figure adapted from Larroude et al., 2018

In *Y. lipolytica* some **genomic tools** have been developed for metabolic engineering (Figure 3), including a disruption cassette system for gene knock-out (Fickers et al., 2003) and transcription activator-like effector nucleases (TALENs) (Rigouin et al., 2017). However, the low efficiency and time-consuming procedures involved in these methods make multiple modifications fastidious. The emergence of the CRISPR-Cas system offers a potential solution for these problems due to its high efficiency and ease to operate.
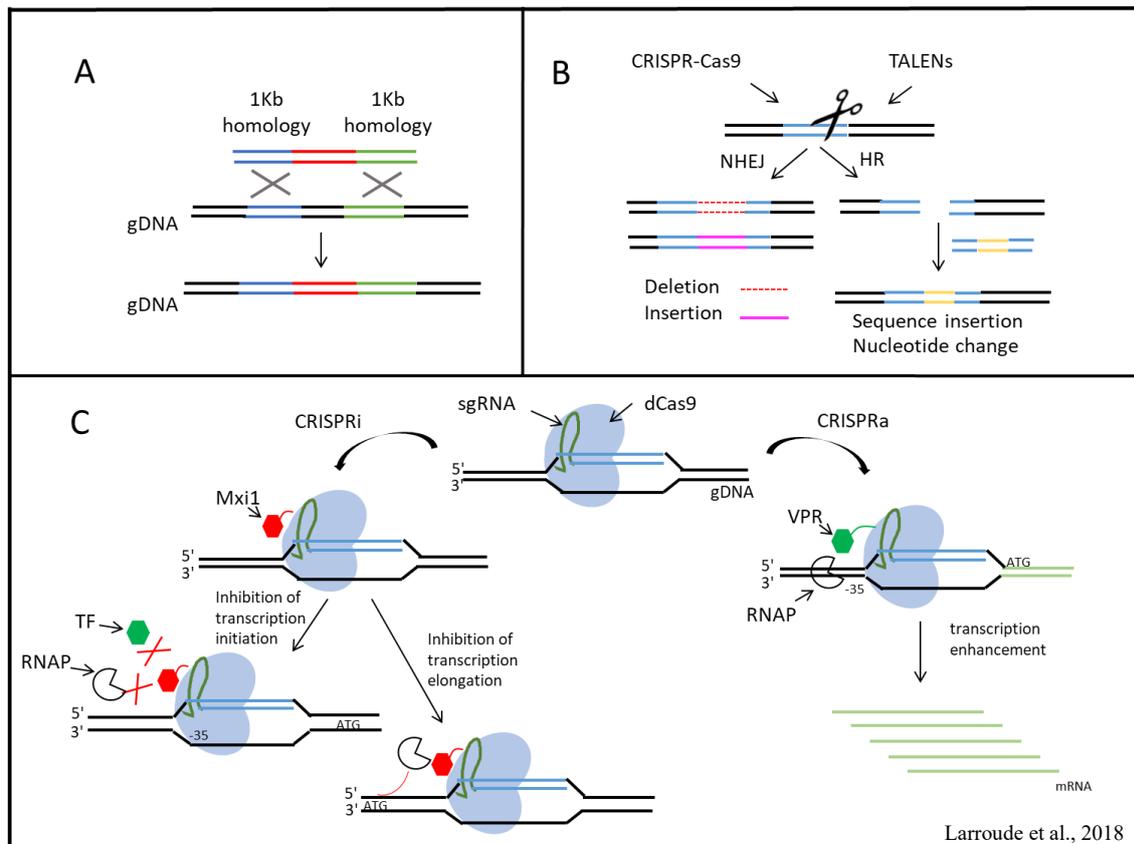
Figure 3. Chromosome editing tools and targeted genome engineering
A. Representation of the homologous recombination (HR) approach, which requires long (~1 kb) homologous flanking fragments to be efficient in *Y. lipolytica*. **B**. TALENs and Cas9 are programmable nucleases that recognize and bind to specific DNA sequences, causing double-strand breaks (DSBs), which induce non-homologous end joining (NHEJ) or HR. NHEJ introduces random insertions and deletions into the genome. Templates with homology arms can be added to take advantage of natural HR mechanisms to either modify single nucleotides or to insert new sequences. It should be noted that Cas9 introduces blunt breaks, while Fok1, the TALEN endonuclease, introduces a staggered cut (for simplicity, this difference is not shown in the figure). **C**. On the left, a CRISPR interference (CRISPRi) system is illustrated. The dCas9-sgRNA complex can either target the promoter inhibiting transcription initiation or target the gene sequence to prevent transcription elongation. On the right, a CRISPR activation (CRISPRa) system is illustrated. dCas9 is fused with a transcription factor and targets the upstream region of the gene, delivering the transcription factor to the promoter; this process enhances transcription efficiency. The abbreviations are as follows: gDNA: genomic DNA; sgRNA: single-guide RNA; dCas9: catalytically inactive Cas9; RNAP: RNA polymerase; TF: transcription factor; Mxi1: repressor; and VPR: synthetic activator domain
Figure adapted from Larroude et al., 2018

The CRISPR-Cas system basically consists of an endonuclease enzyme guided by a RNA structure to a specific DNA sequence that is digested (Jinek et al., 2012). The modifications in the genome are driven by the natural DNA reparation mechanisms of the cell after DNA break, resulting most of the time in randomly sized small deletions or insertions (indels) producing frameshift mutations that lead to the production of short non-sense proteins. If a DNA template is provided, reparation can occur by HR and used to insert a specific DNA sequences at the break point (Figure 3-B). Thus, CRISPR-Cas system turned out to be of great help for targeted genome modification. In addition, it increases the rate of DNA integration by HR in organism where it is highly impaired because of the high NHEJ rate, like in *Y. lipolytica.*

CRISPR-Cas9 system was successfully implemented in *Y. lipolytica* for the first time by Schwartz and colleagues (Schwartz et al., 2016). Cas9 nuclease targets specific genome loci defined by a 20 bp region of single guide RNA (sgRNA) and creates a double-stranded break (DBS) which is then repaired with the subsequent genome modification. The system was rapidly expanded afterwards, allowing gene knock-out/knock-in and repression/activation applications, which was very helpful for accelerating engineering cycles in this yeast (S. Gao et al., 2016; Schwartz et al., 2017b, 2017a; Zhang et al., 2018).

Schwartz and colleagues expressed the gRNA under a synthetic hybrid RNA Polymerase III promoter and used a codon optimized Cas9. Single-gene disruption and HR were more than 90% and 70% effective, respectively. HR efficiency reached 100% when NHEJ was disrupted in the strain (Schwartz et al., 2016). These researchers also managed to integrate multiple genes at different loci and showed that gene integration efficiency depends on the integration site: of the 17 loci tested, only 5 had high CRISPR-Cas9-mediated integration frequencies (48–62%) (Schwartz et al., 2017b). In parallel, a second strategy for CRISPR-Cas9 genome editing in *Y. lipolytica* was developed (S. Gao et al., 2016). It involves expressing a human-codon-optimized Cas9 variant and gRNA flanked by ribozymes under the control of a RNAP II promoter. Its efficiency was 86% after four days of outgrowth. Both systems allow highly effective gene targeting.

Two other CRISPR tools have been developed for use in *Y. lipolytica*. Holkenbrink and colleagues created a toolbox, EasyCloneYALI, containing a set of plasmids for

integrating expression cassettes at a defined genomic locus; users can employ different selection markers or use a marker-free mode. In this study a *ΔKu70* strain was used, making HR more efficient, and Cas9 was integrated into the genome and constitutively expressed. Eleven intergenic sites with high gene expression levels were identified, but only five had efficiencies higher than 80% for marker-free integration (Holkenbrink et al., 2018). Very recently, Gao and colleagues developed a dual CRISPR-Cas9 strategy using paired gRNA to create complete gene knockout via gene excision. Basically, two vectors, each containing a Cas9 gene and a single-guide RNA (sgRNA) cassette, are co-transformed in *Y. lipolytica*. The gRNAs were designed to target areas upstream from the start codon and downstream from the stop codon, which led to complete gene excision when the breaks occurred simultaneously and the resulting genomic regions were end-joined. The strategy was tested on six genes, and excision efficiency reached about 20%. They also used this dual CRISPR-Cas9 strategy to integrate a marker-free DNA fragment into the excision region, reaching an integration efficiency that range from 15% to 37% (Gao et al., 2018).

A CRISPR technique was also developed to control gene expression. In these systems, a catalytically inactive Cas9 (dCas9), able to bind to specific DNA sequences when co-expressed with a gRNA but unable to introduce DSBs, is used to repress or activate gene expression. The repression is achieved by allosteric interference of the initiation or elongation of the transcription by the presence of the Cas9 in the DNA, and can be enhanced by fusing a repressor effector domain, like KRAB or Mxi1. On the other side, the transcription can be activated by fusing a transcriptional activator effector to dCas9 and directing the complex to the promoter region of the gene (Gilbert et al., 2013; Qi et al., 2013) (Figure 3c). Schwartz and colleagues adapted CRISPR interference and activation (CRISPRi and CRISPRa) systems for their use in *Y. lipolytica*. Nine genes were tested for gene repression, and for eight of them at least 50% of transcription was repressed using a multiplex strategy. Repression was enhanced when the Mxi1 repressor, but not the KRAB repressor, was fused with dCas9. This technique was used for the repression of *KU70* and *KU80,* and led to an HR efficiency of 90% (Schwartz et al., 2017a). The same group developed a CRISPRa system to activate genes in *Y. lipolytica*. They screened four different activation domains and several target sites in the promoter region. By adding the VPR activation domain to dCas9 and choosing gRNA targeting locations upstream from the core promoter, they activated

two native β-glucosidases genes, BGLI and BGLII, which allowed growth on cellobiose (Schwartz et al., 2018). In addition, Zhang and colleagues also used a CRISPRi system for gene repression in *Y. lipolytica*—four different repressing constructions (dCpf1, dCas9, dCpf1-KRAB, and dCas9-KRAB) were employed. Gene repression efficiency exceeded 80% when three different sites where simultaneously targeted for the same gene, exploiting a multiplex gRNA strategy. However, no strong repression was achieved by targeting only one point in the genome. As shown by Schwartz and colleagues (2017a), the KRAB domain does not influence dCas9 activity. However, compared to results for dCpf1 alone, the use of dCpf1-KRAB increased repression efficiency by about 30% (Zhang et al., 2018).

All the tools described above were further discussed as part of a larger recent review work on synthetic biology tools for *Y. lipolytica*, published in Biotechnology advances (Larroude et al., 2018). The review, which also includes other tools such as computational tools, is presented at the end of this section.

Since the publication of the mentioned review, further developments were done on genomic modification tools for *Y. lipolytica*, four of which are based on CRISPR system, and they are summarized here after. This is a perfect showcase of the fast expansion of synthetic biology field and on the interest *Y. lipolytica* is getting.

Morse and co-workers developed a CRISPR-Cas9 based on an orthogonal T7 polymerase system for gRNA expression, thus, avoiding host organism-related RNA processing drawbacks when using type II CRISPR-Cas9. The system consists of a T7 polymerase with SV40 nuclear localization tag driven by a strong constitutive promoter and a T7 phi9 promoter driving the guide RNA expression. It was first design in *S. cerevisiae*, were the editing frequency of *CAN1* gene was of around 1% and was in accordance with the results found for the same gene using the commonly used Pol III *snr52* system. The system was then transferred to *Y. lipolytica*, with 60% efficiency for *CAN1* disruption*, and *K. lactis* were *NDT80* was edited with 96% efficiency (Morse et al., 2018).

A recent base editor, Target-AID, designed to recruit cytidine deaminase to the target DNA locus via the CRISPR-Cas9 system was adopted in *Y. lipolytica* for multiplex gene disruption by Bae et al.. The system can directly induce a C to T mutation by deaminating a cytosine ~15-20bp upstream of the PAM. The resulting mismatched U:G

base pair is converted to a U:A base pair by the mismatch repair system and then eventually to a T:A base pair. Single and double gene editing reached 90% and 31% efficiency respectively in Δ*ku70* strains, but no triple mutation was detected (Bae et al., 2019).

Recently, Schwartz and colleagues constructed the first CRISPR-Cas9 genome scale indels library for functional genomics and strain engineering in *Y. lipolytica*. This library contains cells with single knockouts of nearly all genes in the genome. The library was used to identify essential genes and, in accordance with what it is described for other yeasts, they found that around 17 % of the genome was essential. Also, the usefulness of the CRISPR-Cas9 library for novel phenotype screening was demonstrated. They screen for canavanine resistance, which led to the identification of excellent sgRNAs cutters targeting CAN1 as expected, and for increased lipid content in the cells revealing expected and unexpected targets, highlighting the usefulness of genome-scale library screening for identifying non-obvious targets for strain engineering. Furthermore, using a *ΔKu70* strain, making efficient cutting by a particular sgRNA-Cas9 complex a lethal phenotype, they could identify cutting and non- or inefficient cutting gRNA. By doing so, they evidenced that the presence of a polyT motif in a sgRNA correlated with reduced CRISPR-Cas9 activity and that the sgRNAs at the ends of the chromosome were largely inactive, however, the RNA secondary structure did not significantly impact activity (Schwartz et al., 2019).

Yang *et al.* showed very recently for the first time the use of CRISPR-Cpf1 system in *Y. lipolytica* for genome editing (Yang et al., 2019), although this system was already used for inhibiting transcription using an inactive form of the enzyme (Zhang et al., 2018). CRISPR-Cpf1 and CRISPR-Cas9 systems are based on different endonucleases enzymes, thus, their characteristics are different, and they can be considered as complementary systems. For instance, they target different regions of the genome, Cpf1 targets T-rich sequences while Cas9 G-rich zones, and they process the gRNA differently with influence in the system effectiveness (Zetsche et al., 2015). On their work, Yang and co-workers, efficiently disrupted several genes by single and multiplex approaches. When single targeting is used, the indel editing efficiency ranged from 40% to 90% depending on the gene targeted and on the crRNA (RNA guide sequence) used, highlighting, as it is the case for Cas9 systems, the effect of guide sequence on gene-editing efficiency. For multiplex genomic target, their editing efficiency reached more

than 75% for duplex targets and 45% for triplex targets. They also tried several promoters for the transcription and processing of crRNA, and evidenced that both type II and III promoter can be used with variable efficiency depending on the targeted gene (Yang et al., 2019).

In another genome editing approach, a piggyBac transposon system, was successfully transferred into *Y. lipolytica*. The system is based on the expression of a hyperactive TTAA-specific piggyBac transposon from an episomal plasmid and enables insertional mutagenesis and footprint-free excisions. The system allows random integration of DNA into the genome of *Y. lipolytica* at TTAA sequences and was used to create an insertional mutagenesis library which was later used to correlate genomic modifications with phenotypes (Wagner et al., 2018)**.**

Overall, the herein described tools are now extensively used for pathway engineering to exploit *Y. lipolytica* as a cell factory. The impact of these new synthetic biology tools for *Y. lipolytica* is evidenced by the increase in the number of engineered strains and published research works. Even though many more tools are available for the model organisms *S. cerevisiae* and *E. coli*, it is expected that the list of tools dedicated to *Y. lipolytica* will continue to increase.

Research review paper

Synthetic biology tools for engineering *Yarrowia lipolytica*

Larroude M.[a], Rossignol T.[a], Nicaud J.-M.[a], Ledesma-Amaro R.[b],*

[a] *Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France*
[b] *Department of Bioengineering and Imperial College Centre for Synthetic Biology, Imperial College London, London, United Kingdom*

ARTICLE INFO

ABSTRACT

The non-conventional oleaginous yeast *Yarrowia lipolytica* shows great industrial promise. It naturally produces certain compounds of interest but can also artificially generate non-native metabolites, thanks to an engineering process made possible by the significant expansion of a dedicated genetic toolbox. In this review, we present recently developed synthetic biology tools that facilitate the manipulation of *Y. lipolytica*, including 1) DNA assembly techniques, 2) DNA parts for constructing expression cassettes, 3) genome-editing techniques, and 4) computational tools.

1. Introduction

Yarrowia lipolytica is a non-conventional dimorphic yeast with the potential to act as a biotechnological workhorse in a wide range of applications. This organism is often found in fermented foods such as cheese and meat, and it is a good natural producer of certain compounds of industrial interest, including citric acid, erythritol, and various proteins and lipids. The ability of *Y. lipolytica* to grow at high cell densities and to produce large titers of valuable molecules has attracted the attention of the scientific community. Consequently, research carried out with this organism is growing exponentially, which is clearly reflected in the increasing number of related articles and patents. Thanks to the yeast's characteristics and to the development of molecular biology tools specific to it, *Y. lipolytica* has been extensively engineered to produce chemicals and fuels (Cavallo et al., 2017; Guo et al., 2016; Ledesma-Amaro et al., 2016b; Ledesma-Amaro and Nicaud, 2016b).

Synthetic biology is an emerging discipline that aims to apply engineering principles to biological systems to render them more controllable, standardized, and predictable. These latter benefits can significantly accelerate research since new and highly efficient tools can be developed, allowing systems of interest to be constructed more rapidly and at a lower cost. In recent years, different synthetic biology tools have been generated and applied in *Y. lipolytica*, which has further expanded the range of applications for this yeast. *Y. lipolytica* has a metabolism that is well suited to fatty acid production and lipid accumulation and has consequently been used as a host organism for generating large amounts of lipids (Darvishi et al., 2017; Friedlander et al., 2016; Park et al., 2018a, 2018b). More specifically, strains have been engineered to produce large amounts of oleic acid oil with no polyunsaturated fatty acid (Tsakraklides et al., 2018); unusual fatty acids such as hydroxy fatty acids and ricinoleic acid (Beopoulos et al., 2014); oils that resemble cocoa butter, which is rich in stearic acid and could thus be used as an oil substitute in chocolate production (Papanikolaou and Aggelis, 2003a); conjugated fatty acids such as conjugated linoleic acids (Imatoukene et al., 2017; Zhang et al., 2013, 2012); and omega-6 and omega-3 fatty acids like EPA (eicosapentaenoic acid, C20:5) (Xue et al., 2013). The latter process has led to two commercial products by DuPont (Xie et al., 2015).

Other valuable compounds that have been produced include methyl ketones (Hanko et al., 2018); polyhydroxyalkanoates (PHAs), which are good candidates for use in renewable and biodegradable bioplastics (Gao et al., 2015; Haddouche et al., 2011, 2010; Li et al., 2017); carotenoids (Kjaergaard et al., 2017; Larroude et al., 2017); erythritol (Carly et al., 2017); mannitol (Rakicka et al., 2016); and various proteins (Celińska et al., 2018; Dulermo et al., 2017; Madzak, 2015) and organic acids (Markham et al., 2018; Blazeck et al., 2015). *Y. lipolytica* has also been engineered to produced aroma molecules such as 2-phenylethanol (rose-like odor; Celińska et al., 2013) and γ-decalactone (peach-like odor; Braga and Belo, 2016). In addition, we have witnessed the successful development of strategies that decrease bioprocess cost, namely by allowing *Y. lipolytica* to grow on cheap substrates, such as

glycerol, sucrose, starch, inulin, cellobiose, or other waste products available for reuse (Cui et al., 2011; Johnravindar et al., 2018; Lazar et al., 2013; Ledesma-Amaro et al., 2015; Ledesma-Amaro and Nicaud, 2016a; Magdouli et al., 2017; Mirończuk et al., 2016; Spagnuolo et al., 2018; Yang et al., 2015). Strategies have also been developed to facilitate product extraction (Ledesma-Amaro et al., 2016a).

In this review, we describe and assess the most important synthetic biology tools developed to date for *Y. lipolytica*. We focus specifically on DNA assembly techniques, DNA parts for constructing expression cassettes, genome-editing techniques, and computational tools, and we discuss their potential to enhance this yeast's capabilities.

2. DNA assembly techniques

As the cornerstone of synthetic biology, the DNA assembly process allows the construction of novel biological systems and devices using defined, standardized, and well-characterized components. It is a procedure by which multiple DNA fragments are physically linked end to end, creating a target higher-order assembly that is then joined to a vector.

Traditional techniques employing restriction digestion and element-by-element cloning are time consuming and cost inefficient (Celińska and Grajek, 2013; Matthäus et al., 2014). Consequently, significant efforts are being made to develop better cloning strategies and DNA assembly techniques that would allow multigene cassettes to be constructed more quickly and efficiently. It would then be easier to build strains with complex genetic functionalities. These new techniques are also helping to increase the viability and/or transformability of recombinant strains, traits that are often impaired after several rounds of transformation.

In this context, several well-known methodologies have recently been developed for *Y. lipolytica*. Here, we describe the most recent and relevant ones (Fig. 1).

2.1. One-step integration PCR

A simple and cost-effective method developed by Gao et al. (2014) allows the integration of multiple genes (four genes, total size of ~11 kb) by overlap extension PCR (OE-PCR). In their study, even though total efficiency was not very high (~15%), they were able to assemble β-carotene biosynthesis pathways rapidly, within a week's time, by dividing the pathways into four cassettes with ~50 bp overlaps between successive cassettes. The fragments were assembled into a single gene expression cassette that was then used to transform *Y. lipolytica* (Gao et al., 2014).

2.2. Gateway cloning

The Entry/Gateway® method employs site-specific recombination between att sites on interacting molecules to rapidly clone single DNA sequences in multiple destination plasmids (Hartley, 2000). It was recently adapted for use in *Y. lipolytica* by Leplat and colleagues. In their research, Gateway® vectors were combined, at the cloning site, with an overexpression cassette composed of the excisable *URA3* marker, the pTEF promoter, the LIP2 terminator, and zeta sequences, allowing random integration into the *Y. lipolytica* genome. As an example of the technique's utility, a library of alkaline extracellular protease (AEP) overexpression mutants was obtained in a single transformation experiment, using a novel high-throughput transformation method applied using 96-well plates (Leplat et al., 2015). This tool was used to construct more than 150 strains overexpressing individual transcription factors with a view to identifying regulators involved in lipid metabolism (Leplat et al., 2018).

2.3. BioBricks

Wong and colleagues developed a set of BioBrick-based vectors, called YaliBricks, for *Y. lipolytica*. It comprises four compatible restriction enzyme sites (*Avr*II, *Xba*I, *Spe*I, and *Nhe*I) that enable modular genetic engineering and the reuse of parts. Using this system, they were able to characterize 12 endogenous promoters and construct a five-gene biosynthetic pathway for producing violacein within a week's time (Wong et al., 2017). It is a fast and easy method, but it relies on specific restriction-site-free genes.

2.4. Gibson assembly

Gibson assembly allows multiple DNA fragments to be assembled regardless of their length or end compatibility and exploits three different enzymes: exonuclease, DNA polymerase, and DNA ligase (Gibson et al., 2009). Due to its ease of use and flexibility, it has rapidly been adopted as a DNA assembly method in a large range of microorganisms, including *Y. lipolytica*. For instance, Rodriguez and colleagues used the Gibson method to construct vectors to clarify the xylose pathway in *Y. lipolytica* (Rodriguez et al., 2016), while Bhutada and colleagues used it in research demonstrating that the deletion of glycogen synthase has beneficial effects on neutral lipid accumulation (Bhutada et al., 2017).

A similar method, which also uses homolog fragment ends and a single-tube enzymatic reaction, is employed in the commercial In-Fusion Kit (Clontech Laboratories, Inc). The In-Fusion enzymes also generate short regions of single-stranded overlaps between the DNA, facilitating directional assembly.

2.5. Golden gate

The Golden Gate (GG) modular cloning system utilizes type II restriction enzymes (Engler et al., 2008) and establishes a library of standardized and interchangeable DNA parts, which can subsequently be assembled in a single-step, one-pot reaction on a scaffold of predesigned 4 nt overhangs. Recently, a customized GG platform was developed for *Y. lipolytica* (Celińska et al., 2017). It can be used to express one-, two-, or three-customizable transcription units (TUs) in a versatile cassette comprising different genomic integration sites and recyclable auxotrophy markers. This fact means that thirteen elements can be assembled in a very fast and efficient manner. System viability and robustness was validated using a three-TU-bearing cassette encoding carotenoid synthesis genes (Celińska et al., 2017). One of the advantages of GG is that it allows combinatorial assembly, which can be used to efficiently generate libraries. Consequently, a promoter shuffling strategy using GG was used to screen optimum promoter-gene pairs for each transcriptional unit expressed. The best promoter combination was then used to engineer a lipid overproducer strain (Larroude et al., 2017); this research, through a combination of synthetic biology, metabolic engineering, and fed-batch fermentation, achieved the highest production level of β-carotene reported thus far, 6 g/L.

It is important to note that there are no perfect techniques for universal DNA assembly. Each method presents advantages and disadvantages in different situations. Therefore, method selection will depend on one's objectives.

OE-PCR and Gibson assembly are fast and simple, but specific primers are required for each assembly. As a result, these techniques are less versatile if, for example, the goal is to construct combinatorial assembly libraries. Moreover, as these methods are based on DNA annealing and polymerase elongation, it can be hard to correctly assemble hybrid promoters containing several copies of upstream activation sequences (UASs). Another constraint is that the final construct must be checked via sequencing to ensure there are no sequence errors due to polymerase amplification.

In the case of the BioBricks or GG techniques, it is necessary to
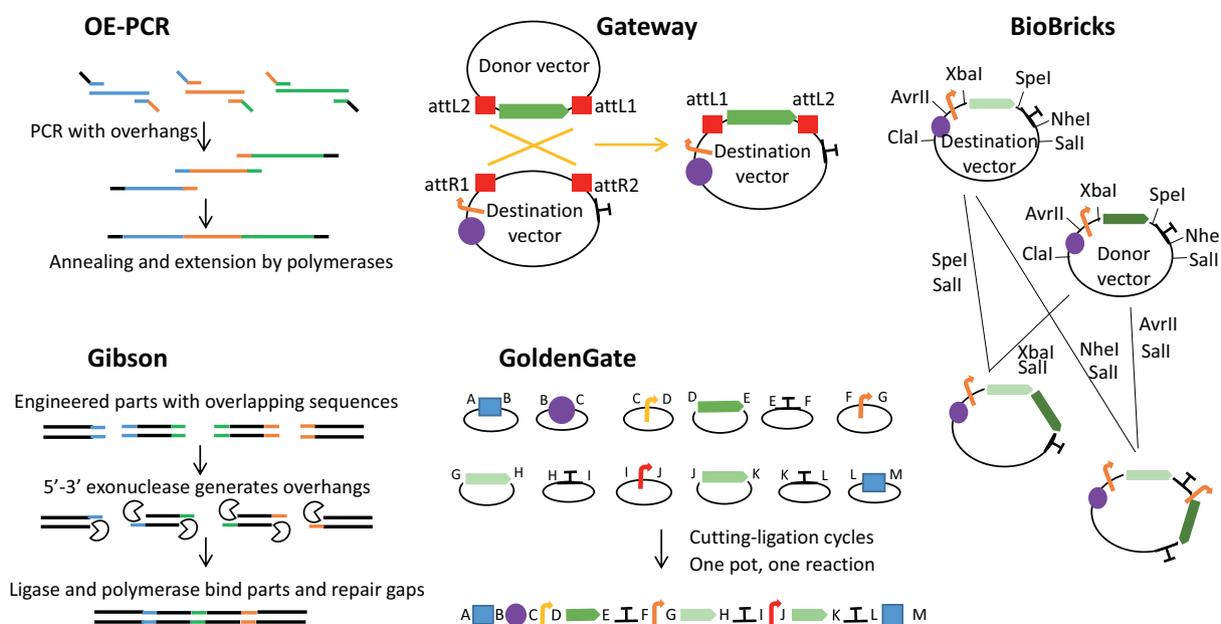
Fig. 1. Summary of DNA assembly techniques.

OE-PCR is a two-step PCR. During the first step, complementary overlapping overhangs are added to the parts to be assembled. During the second step, the parts hybridize with each other and form the new assembly via extension. In the Gateway method, the gene of interest, which has been cloned into the entry vector, is transferred into the destination vector via att site recombination. The expression vector obtained is then digested to release the expression cassette and used to transform Y. lipolytica. The BioBricks technique is used to clone parts via restriction digestion and the subsequent ligation of the resultant compatible sticky ends. YaliBricks vectors were designed to have AvrII, XbaI, SpeI, and NheI endonuclease site recognition. The ligation of the compatible overhangs produces a scar that is no longer recognized by either enzyme, which allows for subsequent assembly steps using more DNA parts. In Gibson assembly, parts are synthesized to overlap by 30+ bp. Their ends are then processed by an exonuclease that creates single-stranded 3′ overhangs, which facilitates annealing. The overhangs are fused together using a polymerase, which fills in gaps within each annealed fragment; a ligase seals gaps in the assembled DNA. Golden Gate assembly exploits type II enzymes, which cut outside their recognition sites to excise parts with arbitrarily defined four-base overhangs. Through the careful selection of compatible overhangs, such parts can be assembled altogether in a defined order. In the figure, the letters A to M represent different compatible 4-nt overhangs; the yellow, orange, and red arrows represent promoters; the green arrows represent genes; the violet circles represent markers; the blue squares represent insertion sequences; and the Ts represent terminators.

construct a library encompassing the different parts to be used in the assembly. Library construction can be long, but, once it is completed, assembly is fast and versatile. There is also no need to verify the final assembly via sequencing (restriction-enzyme-based verification is enough). There are some important differences between the two techniques, however. Compared to GG, the BioBricks method requires to be devoided of greater number of restriction recognition sites. In addition, in GG, full assembly is carried out via a single reaction in a single pot. In contrast, when using BioBricks, parts are assembled one after the other, meaning that vectors should be digested, purified, and ligated at each cloning step. Finally, in GG, a reporter gene (e.g., red fluorescent protein) is used to preliminarily identify clones in which the assembly failed to be incorporated into the destination vector. As a result, GG is faster and easier to use.

The Gateway method does not involve high-throughput assembly technology, but it is very useful because the destination vectors of genes or constructions can be easily changed, which facilitates the functional analysis of genes and protein expression.

3. DNA parts for constructing expression cassettes

Controlling gene expression is a critical component of metabolic engineering, where the expression levels of the different enzymes in the pathway of interest must be balanced to maximize metabolic fluxes and minimize protein synthesis costs. Protein expression can be controlled by different means, including (but not limited to) transcription, mRNA stability, translation efficiency, or protein stability. Here, we discuss the most important tools that have been developed in *Y. lipolytica* to control the expression of genes, proteins, and other DNA parts that allow the

creation of efficient integration cassettes and plasmids (Table 1).

3.1. Promoters

In eukaryotic synthetic biology, selecting promoters according to their strength is the most widespread method for controlling gene expression in metabolic engineering techniques. However, it is important to note that higher protein expression is not always obtained using stronger promoters (Dulermo et al., 2017). Therefore, studies exploring promoter characteristics and engineering are needed to be able to consistently modify gene expression, which is a key parameter to be optimized in pathway engineering. Consequently, significant efforts have been made to develop promoters exhibiting a wide range of transcriptional activities.

In *Y. lipolytica*, the first strong promoters to be isolated and characterized were 1) the promoter from the *XPR2* (pXPR2) gene, which codes for an alkaline extracellular protease (Ogrydziak and Scharf, 1982), and 2) the constitutive promoter of *TEF* (pTEF), which codes for translation elongation factor-1 (Müller et al., 1998). Other native promoters that have been characterized or used to express heterologous genes in *Y. lipolytica* are pTDH1, pGPM1 (Hong et al., 2012), pEXP1, pFBAINm, pGPAT, pGPD, and pYAT (Xue et al., 2013). As mentioned earlier, Wong et al. (2017) selected and characterized eleven endogenous promoters, primarily associated with lipogenic pathways, in addition to the well-described pTEF promoter. pTEF was the most active promoter, followed by pGAP, pICL, and pACL2. The other promoters tested were pDGA1, pACC, pIDH2, pFAS2, pFAS1, pPOX4, pZWF1, and pIDP2.

To increase the current strength of available promoters, hybrid

Table 1
DNA parts used in expression cassette construction.

DNA Part	Characteristics	References
Promoters		
pTEF	constitutive; native	Muller et al., 1998
pTDH1, pGPM1, pFBA$_{IN}$	constitutive; native	Hong et al., 2012
pEXP1, pGPAT, pGPD	constitutive; native	Xue et al., 2013
pGAP, pACL2	inducers were not determined; native	Wong et al., 2017
pXPR2	inducible by peptones; native	Ogrydziak and Scharf, 1982
pPOX2	inducible by fatty acids and alkanes; repressed by glucose and glycerol; native	Juretzek et al., 2000
pPOT1	inducible by fatty acids and alkanes; repressed by glucose and glycerol; native	Juretzek et al., 2000
pLIP2	inducible by fatty acids and alkanes; native	Sassi et al., 2016
pICL	inducible by ethanol, fatty acids, and alkanes; native	Juretzek et al., 2000
pYAT1	induced by nitrogen-limited conditions; native	Xue and Zhu, 2012
hp4d	hybrid promoter derived from pXPR2; contains four copies of UAS1XPR2 fused upstream from a minimal core LEU2 promoter; growth phase dependent; hybrid	Madzak et al., 2000
n UAS1$_{XPR2}$-LEU	hybrid promoter derived from pXPR2; core minimal LEU2 promoter; n = number of UASs (up to 32); hybrid	Blazeck et al., 2011
n UAS1$_{XPR2}$-TEF	hybrid promoter derived from pXPR2; core minimal TEF promoter; n = number of UASs (up to 16); hybrid	Blazeck et al., 2011
pEYK1, pEYD1	strongly induced by erythritol and erythrulose; repressed by glucose and glycerol; native and hybrid (p3AB-EYK)	Trassaert et al., 2017 Park et al., unpublished
Terminators		
XPRt	native terminator sequence	Franke et al., 1988
Lip2t	native terminator sequence	Pignede et al., 2000
Minimal XPRt	100-bp non-coding 3′ sequence	Swennen et al., 2002
CYC-t	*S. cerevisiae* terminator sequence	Blazeck et al., 2011; Mumberg et al., 1995
Synthetic	short synthetic terminators	Curran et al., 2015
Markers		
LEU2, URA3, Lys5	auxotrophy complementation	Barth and Gaillardin, 1996
Ura3d4	promoter-defective gene; several copies needed to restore auxotrophy	Le Dall et al., 1994
SUC2	gene from *S. cerevisiae*; sucrose utilization	Nicaud et al., 1989
EYK1	gene from *Y. lipolytica*; erythritol as carbon source	Vandermies et al., 2017
ptxD	gene from *Pseudomonas stutzeri*; growth in phosphite-containing media	Shaw et al., 2016
Hph	gene from *E. coli*; resistance to Hygromycin B	Otero et al., 1996; Tsakraklides et al., 2018
nat1	gene from *Streptomyces noursei*; resistance to nourseothricin	Kretzschmar et al., 2013; Tsakraklides et al., 2018
guaB	gene from *E. coli*; resistance to mycophenolic acid	Wagner et al., 2018
ylAHAS W572 L	*Y. lipolytica* mutant resistant to chlorimuron ethyl herbicide	Wagner et al., 2018
Tags		
tripeptide AKI or SKL	peroxisomal targeting signal	Xue et al., 2013
GPI anchor domains	signal for surface display; covalent bonds with cell wall β-1,6 glucans	Yue et al., 2008; Yuzbasheva et al., 2011; Moon et al., 2013
Flocculation domains	signals for surface display; non-covalent bonds with cell surface mannan chains	Yang et al., 2009
CBM	signals for surface display; non-covalent bonds with chitin	Duquesne et al., 2014
Pir	signals for surface display; covalent bonds with β-1,3 glucans	Duquesne et al., 2014
Oleosin C-t domain	targets lipid bodies	Han et al., 2013
Synthetic consensus secretory signal	MKFSAALLTAALA(S:V)AAAAA	Celińska et al., 2018
Fluorescent tag	reveals expression and localization	Bredeweg et al., 2017
Elements for exogenous DNA maintenance		
1 kb homologous flanking fragments	required for DNA integration by HR	Barth and Gaillardin, 1996
ZETA elements	integration at a zeta locus in zeta-containing strains; random integration in zeta-free strains	Pignede et al., 2000
rDNA	repeated genomic sequences	Le Dall et al., 1994
ARS68/ARS18	for maintenance of autonomously replicating vectors	Matsuoka et al., 1993; Fournier et al., 1993

GPI: glycosyl phosphatidyl inositol, CBM: chitin-binding module, Pir: protein internal repeat.

promoters have been developed. The functional dissection of pXPR allowed one of its UASs (UAS1$_{XPR2}$) to be identified. Madzak and colleagues developed hybrid promoters containing up to four copies of UAS1$_{XPR2}$ fused upstream from a core minimal LEU2 promoter, which were named the hp1d, hp2d, hp3d, and hp4d promoters, respectively (Madzak et al., 2000). These four strong hybrid chimera promoters resulted in a linear increase in promoter strength as a function of the number of tandem UAS1$_{XPR2}$ elements (Madzak et al., 2000). The hp4d promoter is a widely used tool for heterologous gene expression in *Y. lipolytica*.

Since then, other hybrid promoters have been developed using this basic approach of associating multiple UAS tandem elements with a core promoter. Hybrid promoters containing up to 32 copies of UAS1$_{XPR2}$ upstream from the minimal LEU2 core promoter and up to 16 copies of UAS1$_{XPR2}$ upstream from the TEF core promoter have also been constructed. Some can increase expression efficiency eight-fold, compared to the known endogenous promoter in *Y. lipolytica* (Blazeck et al., 2011).

A general strategy for efficiently building synthetic promoters de novo is to both increase native expression capacity and to produce libraries for customizing gene expression; such approaches have been described by several groups (Blazeck et al., 2013, 2011; Dulermo et al.,

2017; Shabbir Hussain et al., 2016; Trassaert et al., 2017). They identified novel UAS and core promoters, via promoter truncation and fragment dissection analysis, and created libraries of characterized strong hybrid promoters by fusing activating regions and core regions that were treated as independent synthetic parts. They identified UAS elements in the TEF promoter as well as the corresponding core minimal TEF promoter and then demonstrated that the ability of a UAS element to amplify expression is independent of the core promoter element. However, the magnitude of the amplification does depend on the core. Thus, the choice of both elements helps determine hybrid promoter strength, implying that it would be possible to design hybrid promoters displaying specific expression strengths. In other recent research (Shabbir Hussain et al., 2016), promoter strength was investigated by shuffling promoter constitutive elements (UAS, proximal promoter, TATA box, and core promoter); the results showed that gene expression can be fine-tuned by engineering such elements.

Another strategy used to enhance expression levels involves retaining an upstream intron with its corresponding promoter. Even though this intron-mediated enhancement (IME) has been observed in several organisms, including both plants and mammals, the mechanism remains elusive and is not effective for all introns (Gallegos and Rose, 2015). In *Y. lipolytica*, 15% of genes contain an intron (Neuvéglise et al., 2011), which means IME may play an important role in regulating gene expression in this organism. Hong and colleagues used pFBA with the native FBA intron (FBA1$_{IN}$) and found that its strength was five times that of pFBA on its own. The efficacy of IME was further confirmed when a chimeric promoter, GPM1::FBA1, containing the 5′-region of the *FBA1* gene was attached to pGPM1 (Hong et al., 2012). The results of this study were similar to those previously described by Juretzek and colleagues, who discovered that expression levels climbed two-fold when pG3P contained the intron of *G3P* (pG3PB2), as compared to a promoter without an associated intron (Juretzek et al., 2000). The TEF$_{IN}$ promoter has also been shown to increase gene expression levels 17-fold, compared to the constitutive TEF promoter (Tai and Stephanopoulos, 2013).

Inducible promoter systems offer the advantage of being able to control gene expression levels based on the presence of specific inducer or repressor molecules. This feature is desirable, for example, in fermentation processes because it allows the uncoupling of growth and production phases. The first of such promoters to be described were induced by hydrophobic substrates such as alkanes and lipids. These promoters are mainly encoded by the peroxisomal acyl-CoA oxidase 2 (POX2) gene, the peroxisomal 3-ketoacyl-thiolase (POT1) gene, the extracellular lipase Lip2 (LIP2) gene, and the cytochrome P450 oxidase (ALK1) gene (Juretzek et al., 2000; Sassi et al., 2016).

Recently, Trassaert and colleagues (Trassaert et al., 2017) isolated, characterized, and modified the promoter of the *EYK1* gene (pEYK1), which is induced by erythritol and erythrulose and repressed by glucose and glycerol. They showed that a hybrid promoter containing two additional tandem copies of the short (48-bp) UAS1$_{EYK1}$ located upstream from the EYK1 promoter resulted in a 3.3-fold increase in expression. Deletion of the *EYK1* gene further improved expression levels: by preventing the catabolism of erythritol and erythrulose, induction was enhanced.

They are no complete report on the comparison of all available promoters. However, Darvishi et al. (2018) recently compared the relative strength of different promoters. We do know that promoters vary in size. Generally, native promoters range between 700 and 1000 bp in length, while synthetic promoters can exceed 1000 bp, depending on the number of repetitions they contain. That said, the recently described pEYK that includes regulatory elements useful in gene expression is only 300 bp long (Trassaert et al., 2017). The identification of *cis*-regulatory modules (CRMs) can help construct very short promoters, which can reduce assembly size and diminish the likelihood of homologous recombination due to repeated sequences.

3.2. Terminators

Terminators are essential for completing the transcription process because they affect mRNA stability and half-life and thus influence net protein synthesis (Geisberg et al., 2014; Mischo and Proudfoot, 2013). Native *S. cerevisiae* terminators have been successfully used not only in *Y. lipolytica* but also in other yeasts, suggesting a high degree of transferability across species (Wagner and Alper, 2016). Curran and colleagues recently evaluated a subset of short synthetic *S. cerevisiae* terminators in *Y. lipolytica* and found that GFP fluorescence was 60% greater, compared to results obtained using the wild-type CYC1 terminator. However, mRNA output was lower than in *S. cerevisiae*, indicating that there may still be some undefined organism-specific factors involved in termination (Curran et al., 2015).

Despite their importance, terminators have been studied in less detail than promoters, both in yeasts in general and in *Y. lipolytica* in particular. Creating de novo synthetic terminators would have benefits, such as minimizing the risk of undesirable homologous recombination. Also, shorter sequences could be used that have the same net effects as native terminators. Terminator studies are a promising area of research in *Y. lipolytica*.

3.3. Tags (secretion, localisation, and visualisation)

Different kinds of tags can be attached to homologous or heterologous proteins to direct them into different cellular or extracellular spaces, which is helpful for purification purposes or to compartmentalize pathway reactions. Some tags can also be used to determine protein localization or expression and thus partially or fully characterize a specific pathway. Here, we describe the tags most commonly used in *Y. lipolytica*.

Heterologous proteins can be tagged for 1) release into the cultivation medium, 2) display on the cell surface, or 3) incorporation into target intracellular organelles; fusion with a proper targeting sequence is required. For efficient secretion, a signal sequence is fused upstream from the mature sequence of the protein of interest. The most commonly used secretion signals are derived from the *Y. lipolytica XPR2* gene (which encodes the extracellulare protease AEP) or *LIP2* genes (which encode the extracellular lipase Lip2). However, other heterologous signal peptides (SPs) have also been used successfully (see reviews by Madzak, 2015; Madzak and Beckerich, 2013). Several anchor signals leading to surface display have also been developed, such as the glycosylphosphatidylinositol (GPI) anchor domain, which creates covalent bonds with cell wall β-1,6 glucans (Moon et al., 2013; Yue et al., 2008; Yuzbasheva et al., 2011); flocculation domains, which non-covalently bind to cell-surface mannan chains (Yang et al., 2009); chitin-binding modules (CBMs), which non-covalently bind to chitin in the cell wall (Duquesne et al., 2014), and the protein-internal-repeat (Pir) domain, which covalently binds to β-1,3 glucans (Duquesne et al., 2014). These anchor signals can be used in a variety of applications, such as bioconversion, biosensing, or high-throughput screening of enzymatic activity. Among the intracellular organelles, *Y. lipolytica* peroxisomes can be targeted by the peroxisomal targeting signal (PTS) domain, as the tripeptides AKI or SKL can be fused to the C-terminus of the protein (Haddouche et al., 2010; Xue et al., 2013). *Y. lipolytica* lipid bodies can be targeted by fusing the protein to the C-terminal domain of oleosin, a structural protein embedded in the phospholipid membrane of plant oleosomes (Han et al., 2013). The nucleus can be targeted by fusing the well-known viral SV40 nuclear localization sequence (NLS) (PKKKRKV) to the protein (e.g., for efficiently targeting Cas9; Schwartz et al., 2016). Other putative nuclear localization signals (Campos-Góngora et al., 2013) as well as putative mitochondrial targeting sequences (Bakkaiova et al., 2014; Kerscher et al., 2004) have been identified for *Y. lipolytica* proteins. However, no further analyses were done, to our knowledge, to characterize these endogenous targeting sequences. Systematic research focused on the identification and

characterization of these endogenous targeting sequences is lacking.

Very recently, Celińska et al. (2018) analyzed the potential of 10 different SPs to facilitate the secretion of two heterologous proteins in *Y. lipolytica*. The study examined both previously described and novel SPs. The latter were identified via genomic DNA data mining and are native secretory proteins that are highly expressed in *Y. lipolytica*. Secretory capacity was assessed experimentally and compared with that obtained with known secretory tags. The most potent SPs turned out to be the novel SP1, SP3, and SP4 (from proteins encoded by YA-LI0B03564g, YALI0E22374g, and YALI0D06039g, respectively). The researchers also suggested a consensus sequence (MKFSAALLTA-ALA(S:V)AAAAA) for a potentially robust synthetic SP, which could be used to expand the molecular toolbox for engineering *Y. lipolytica*.

Using fluorescent proteins for characterizing cellular localization and expression is a well-known and widely used technique. In *Y. lipolytica*, this technique has been used to study and characterize transcription factors (Martinez-Vazquez et al., 2013) and hybrid promoters (Blazeck et al., 2013, 2011; Dulermo et al., 2017; Shabbir Hussain et al., 2016; Trassaert et al., 2017). A multipurpose vector for rapidly expressing fluorescently tagged proteins in *Y. lipolytica* was recently developed, which has streamlined analysis of protein localization (Bredeweg et al., 2017). In the latter study, the authors described the localization of enzymes involved in lipid synthesis. They also generated an atlas of strains with green fluorescent organelles, by tagging genes with GFP at their endogenous locus. These intracellular markers respond to native promoter controls and are non-essential proteins displaying consistently high levels of expression. This tool may facilitate cell biology research on organelles and on the colocalization of biosynthetic enzymes and pathways. Using green fluorescent protein (GFP) to tag organelle-specific proteins can overcome many of the limitations associated with dyes such as DAPI, Nile Red, or ER tracers, which include toxicity, poor penetration, and variability due to growth conditions, age, or nutrient availability (Bredeweg et al., 2017).

3.4. Plasmid vectors and genomic integration cassettes

In *Y. lipolytica*, like in other yeasts, the transforming vectors are shuttle vectors-hybrids between yeast- and bacteria-derived sequences. The bacterial component consists of a replication origin and a bacterial marker gene from *E. coli*. For expression in *Y. lipolytica*, a selection marker, transcriptional unit, and maintenance elements are needed. Two types of shuttle vectors, differing in their mode of maintenance in yeast cells, can be used: (i) episomal vectors (replicative) and (ii) integrative vectors (designed to be integrated into the yeast chromosome). As no natural episome has ever been detected in *Y. lipolytica*, replicative plasmids have been designed using chromosomal autonomously replicating sequence/centromere (ARS/CEN) replication origins (Fournier et al., 1993; Matsuoka et al., 1993). However, for the purposes of heterologous expression and/or genetic engineering, the use of such vectors is limited in *Y. lipolytica* because of low copy numbers (~1–3 plasmids/cell) and the high frequency of loss. That said, a two- to six-fold increase in gene expression was obtained using an expression cassette cloned into a replicative vector, as compared to the same cassette integrated into the genome (Nicaud et al., 1991). Replicative plasmids are of great interest for transient protein expression, like when Cre recombinase is produced for marker excision (Fickers et al., 2003) and Cas9 is produced for genome editing (Schwartz et al., 2016). Liu et al. (2014) engineered a replicative vector by fusing a promoter upstream from the CEN element. Although an 80% increase in plasmid copy number resulted, biased plasmid segregation was also observed.

Consequently, integrative vectors remain the vectors of choice (Madzak et al., 2000; Nicaud et al., 2002) for two main reasons. First, they are extremely stable. Second, they can be used to carry out multiple integration, with its correlated increase in gene expression. Multicopy integration can be achieved in *Y. lipolytica* by targeting repeated sequences (see below) or by using a defective marker, which allows

more than 30 copies to be incorporated into the genome (Juretzek et al., 2001; Nicaud et al., 2002).

Integration into the genome can be achieved at specific target sites via homologous recombination. The process requires large (0.5–1 kb) and homologous 5′ and 3′ flanking regions. Integration can also take place within repeated genomic sequences, such as the rDNA region (Juretzek et al., 2001; Le Dall et al., 1994), and the long terminal repeated sequence of *Y. lipolytica*'s retrotransposon, Ylt1, called the zeta region (714 bp) (Bordes et al., 2007; Juretzek et al., 2001; Nicaud, 2012; Pignede et al., 2000). Within the *Y. lipolytica* genome, around 200 copies of rDNA sequences are present; they are located on every chromosome (Casaregola et al., 1997) and have been used for multiple gene integration (Bulani et al., 2012; Celinska et al., 2016; Juretzek et al., 2001; Le Dall et al., 1994). Interestingly, when the zeta sequences on both sides of the integration cassette are used, insertion mainly occurs at a zeta locus in zeta-containing strains, while insertion is random in non-zeta-containing strains (Pignede et al., 2000). By taking advantage of the absence of Ylt1 in some strains, a specific locus integration platform has been developed that uses the zeta sequence to perform targeted integration at the *ura3* locus via a single cross-over event and homologous recombination (Bordes et al., 2007; Juretzek et al., 2001). Holkenbrink et al. (2018) identified 11 intergenic sites to be targeted using specific integrative expression vectors. These loci were selected with a view to promoting high gene expression levels and limiting the effects of expression cassette integration on growth.

To preclude the presence of bacterial DNA in the yeast strain after transformation, which can be a serious drawback, especially in commercial applications, "auto-cloning" expression vectors were developed (Pignede et al., 2000), from which the bacterial moiety can be removed prior to transformation by restriction digestion and agarose gel electrophoresis. The purified yeast expression cassettes are generally composed of an auxotrophy marker, the gene of interest under a specific promoter and terminator, and sequences for targeted integration into the genome; they can be used by themselves to transform the recipient strain.

3.5. Selection markers

Both genome integration and plasmid maintenance modification rely on selection markers. In *Y. lipolytica*, both auxotrophy and dominant markers are available. Auxotrophy markers, which can only be used with specific strains, remain the best choice for performing selection in *Y. lipolytica* (e.g., leucine, uracil, lysine, or adenine) (Barth and Gaillardin, 1996). A particular auxotrophy marker of note is *ura3d4*, a promoter-defective version of the *URA3* marker that is unable to correct auxotrophy when present in a single copy; thus, as mentioned before, it can be used to achieve multiple integration (Le Dall et al., 1994).

Dominant markers tend to be more broadly employed. Those that are available for *Y. lipolytica* include the *E. coli hph* gene (conferring hygromycin resistance) (Otero et al., 1996), the *Streptomyces noursei nat1* gene (conferring nourseothricin resistance) (Kretzschmar et al., 2013), the *Y. lipolytica* AHAS W572 L mutant (conferring chlorimuron ethyl herbicide resistance), the *E. coli guaB* gene (conferring mycophenolic acid resistance) (Wagner et al., 2018), and the *Streptoalloteichus hindustanus ble* gene (conferring zeocin resistance) (Tsakraklides et al., 2018). Other dominant markers involve the utilization of a specific carbon source, such as sucrose in the case of *S. cerevisiae* invertase expression (*SUC2* gene) (Nicaud et al., 1989a, 1989b, Lazar et al., 2013) or erythritol in the case of erythrulose kinase expression (*EYK1* gene) in a *Y. lipolytica* strain lacking this gene (Vandermies et al., 2017). However, use of these markers has proven to be difficult because of residual growth on sucrose impurities and the high level of spontaneous resistance in transformed cells, respectively (Barth and Gaillardin, 1996).

Shaw et al. (2016) engineered *Y. lipolytica* to exploit a naturally rare

compound, potassium phosphite, and then turn around to supply phosphorus in a phosphate-deficient medium. The expression of phosphite dehydrogenase (*ptxD* gene) from *Pseudomonas stutzeri* can be used as a marker since it allows *Y. lipolytica* to grow in a phosphite-containing medium, which wild-type strains cannot do. The use of such rare compounds as source of nutrients can also prevent the growth of undesirable and/or foreign organisms. This competitive advantage can be of key importance in preventing contamination in large-scale fermentation operations; with minimal costs and the spread of antibiotic resistance genes is limited.

When creating strains harboring several integration cassettes, marker availability can be limiting. To solve this problem, Fickers and colleagues designed an approach combining the sticky-end polymerase chain reaction (SEP) method and the Cre-lox recombination system to facilitate efficient marker rescue and reuse. Upon expression of Cre recombinase, the marker was excised at a frequency of 98%, via recombination between the two lox sites. This method was shown to be very helpful in carrying out multiple gene deletions over a short period of time in *Y. lipolytica* (Fickers et al., 2003).

4. Genome-editing techniques

Industrial-scale fermentation processes require that *Y. lipolytica* strains display a high degree of genetic stability; this stability is attained by incorporating genetic modifications into the chromosomes. When carrying out DNA repair, *Y. lipolytica* preferentially uses non-homologous end joining (NHEJ), as opposed to homologous recombination (HR) (Richard et al., 2005). This fact explains why long (~1 kb) homologous flanking fragments are required for HR (Fig. 2.a); however, their frequency of proper integration remains low. The *ku70* gene encodes a DNA-binding protein responsible for double-strand break repair during NHEJ. Its disruption significantly hinders NHEJ efficiency and increases the use of HR (Kretzschmar et al., 2013; Verbeke et al., 2013). Therefore, Δ*ku70* strains are commonly used for targeted gene insertion in *Y. lipolytica*.

As mentioned before, Fickers et al. (2003) designed a knock-out system for performing multiple gene deletions over a short period of time in *Y. lipolytica*. This process was achieved by constructing disruption cassettes harboring 1) regions homologous to the promoter and gene terminator regions intended for deletion; 2) an excisable marker with lox sequences on either side; and 3) a Cre recombinase that facilitates efficient marker rescue and reuse. Disruption cassette construction was later improved by using asymmetric *Sfi*I sites instead of *I-Sce*I, which simplified and speeded up assembly of cassette elements (Vandermies et al., 2017).

More recently, engineered nucleases that cleave specific DNA sequences in vivo have been developed for targeted mutagenesis. These nucleases enable efficient and precise genetic modifications to be carried out by inducing DNA double-stranded breaks (DSBs), which trigger DNA repair mechanisms that ultimately result in endogenous gene editing. The most widely used enzymes across different microorganisms are zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and Cas9 (Gaj et al., 2013); the latter two were recently developed for use in *Y. lipolytica*.

TALENs were created by fusing transcription activator-like effectors (TALEs) to the catalytic domain of the *Fok*I endonuclease. By customizing the TALE DNA binding domain, the DNA DSBs can be directed to occur at a specific target site (Christian et al., 2010) (Fig. 2.b). This technique was recently applied in *Y. lipolytica* to generate fatty acid synthase (FAS) mutants and proved to be very efficient in inducing targeted genome modifications: mutants were generated via error-prone NHEJ repair at the targeted locus in 97% of transformants. When homologous exogenous DNA was added to the TALEN targeted site, HR-mediated repair occurred in 40% of clones. This technique was used to directly produce site-directed mutagenesis in the *Y. lipolytica* genome (Rigouin et al., 2017).

A clustered regularly interspaced short palindromic repeats (CRISPR) technique can also be used. The CRISPR-Cas9 system consists of a Cas9-targeted nuclease that can be programmed with guide RNA (gRNA) to generate DSBs at specific DNA sites (Jinek et al., 2012) (Fig. 2.b). Recently, the CRISPR-Cas9 system from *Streptococcus pyogenes* was adapted to perform marker-free gene disruption and integration in *Y. lipolytica* (Schwartz et al., 2016). Schwartz and colleagues expressed the gRNA under a synthetic RNAP III promoter, and Cas9 was codon optimized for *Y. lipolytica*. Single-gene disruption and HR were more than 90% and 70% effective, respectively, when Cas9 and the gRNA were cotransformed using donor DNA. HR efficiency reached 100% when NHEJ was disrupted in the strain (Schwartz et al., 2016). Schwartz and colleagues also managed to integrate multiple genes at different loci without resorting to marker recovery. Nevertheless, gene integration efficiency depends on the integration site: of the 17 loci tested, 5 had high CRISPR-Cas9-mediated integration frequencies (48–62%) (Schwartz et al., 2017b). Around the same time, a second strategy for CRISPR-Cas9 genome editing in *Y. lipolytica* was developed (Gao et al., 2016). It involves expressing a human-codon-optimized Cas9 variant and gRNA flanked by ribozymes under the control of a RNAP II promoter; efficiency was 86% after four days of outgrowth. Both systems allow highly effective gene targeting.

Two other CRISPR tools have been developed for use in *Y. lipolytica*. Holkenbrink and colleagues created a toolbox, EasyCloneYALI, for easily performing genome editing in *Y. lipolytica* via CRISPR-cas9 technology—the standardized promoters, genes, and plasmids can be reused and easily exchanged (Holkenbrink et al., 2018). The researchers constructed a set of plasmids for integrating expression cassettes at a defined genomic locus; users can employ different selection markers or for the marker-free mode. Ku70p had been deleted in this study, making HR more efficient, and Cas9 was integrated into the genome and constitutively expressed. Eleven intergenic sites with high gene expression levels were identified, but only five had efficiencies higher than 80% for marker-free integration. Very recently, Gao et al. (2018) developed a dual CRISPR-cas9 strategy using paired gRNA to create complete gene knockout via gene excision. Basically, two vectors, each containing a Cas9 gene and a single-guide RNA (sgRNA) cassette, are cotransformed in *Y. lipolytica*. The gRNAs were designed to target areas upstream from the start codon and downstream from the stop codon, which led to complete gene excision when the breaks occurred simultaneously and the resulting genomic regions were end-joined. The strategy was tested on six genes, and excision efficiency reached about 20%. The researchers also used this dual CRISPR-cas9 strategy to incorporate donor DNA into the excision region using marker-free integration (i.e., the integrated cassette had no selection marker). Then, a single vector containing the Cas9 gene and the two sgRNAs was constructed, and cotransformation occurred with another vector containing the donor cassette. Integration efficiency ranged from 15% to 37%, depending on the method (HR or homology-mediated end-joining, respectively).

A CRISPR technique was also developed for controlling gene expression. Schwartz and colleagues adapted CRISPR interference and activation (CRISPRi and CRISPRa) systems for use in *Y. lipolytica*. In these systems, a catalytically inactive Cas9 (dCas9), which is able to bind to DNA complementary to the gRNA spacer sequence but unable to introduce DSBs, targets the promoter region of a gene of interest, repressing or activating transcription, respectively (Fig. 2.c). First, the researchers showed that the system functioned for gene repression (Schwartz et al., 2017a)—for 8 of the 9 genes tested, at least 50% of transcription was repressed using a multiplex strategy. Repression was enhanced when the Mxi1 repressor, but not the KRAB repressor, was fused with dCas9. Finally, the repression of *KU70* and *KU80* led to an HR efficiency of 90%. Later, this same group developed a CRISPRa system to activate genes in *Y. lipolytica* (Schwartz et al., 2018). They screened four different activation domains and several target sites in the promoter region. By adding the VPR activation domain to dCas9 and choosing gRNA targeting locations upstream from the core

Fig. 2. Chromosome editing tools and targeted genome engineering.
A. Representation of the homologous recombination (HR) approach, which requires long (~1 kb) homologous flanking fragments to be efficient in *Y. lipolytica*. B. TALENs and Cas9 are programmable nucleases that recognize and bind to specific DNA sequences, causing double-strand breaks (DSBs), which induce non-homologous end joining (NHEJ) or HR. NHEJ introduces random insertions and deletions into the genome. Templates with homology arms can be added to take advantage of natural HR mechanisms to either modify single nucleotides or to insert new sequences. It should be noted that Cas9 introduces blunt breaks, while Fok1, the TALEN endonuclease, introduces a staggered cut (for simplicity, this difference is not shown in the figure). C. On the left, a CRISPR interference (CRISPRi) system is illustrated. The dCas9-sgRNA complex can either target the promoter inhibiting transcription initiation or target the gene sequence to prevent transcription elongation. On the right, a CRISPR activation (CRISPRa) system is illustrated. dCas9 is fused with a transcription factor and targets the upstream region of the gene, delivering the transcription factor to the promoter; this process enhances transcription efficiency. The abbreviations are as follows: gDNA: genomic DNA; sgRNA: single-guide RNA; dCas9: catalytically inactive Cas9; RNAP: RNA polymerase; TF: transcription factor; Mxi1: repressor; and VPR: synthetic activator domain (Schwartz et al., 2018).

promoter, they activated two native β-glucosidases genes, *BglI* and *BglII*, which allowed growth on cellobiose. Indeed, Guo and colleagues engineered a *Y. lipolytica* strain capable of degrading cellobiose, thanks to the overexpression of these two endogenous genes (Guo et al., 2015).

Zhang et al. (2018) used a CRISPRi system for repressing genes in *Y. lipolytica*—four different repressors (dCpf1, dCas9, dCpf1-KRAB, and dCas9-KRAB) were employed. As it was difficult to achieve strong repression levels with a single gRNA element and identify effective target sites, the group exploited a multiplex gRNA strategy. Gene repression efficiency exceeded 80% when the *gfp* gene was targeted at three different sites. As shown by Schwartz et al. (2017a), the KRAB domain does not influence dCas9 activity; however, compared to results for dCpf1 alone, the use of dCpf1-KRAB increased repression efficiency by about 30%.

The development of the TALEN and CRISPR-Cas9 systems was an important step in modern genomic engineering. Due to their simplicity, efficiency, and affordability, they have both become key genome-editing tools. However, the CRISPR-cas9 system has some advantages: compared to the TALEN system, its molecular tool is much easier to produce, and the technique is also suitable for multiplex genome editing. That said, the fact that TALENs cause breaks only upon dimerization of the *Fok*I domain increases system specificity and reduces

the risk of off-target effects. Regarding the use of the CRISPR-Cas9 system to carry out multiple disruptions, progress must still be made in *Y. lipolytica* to reach efficiency levels comparable to those in *S. cerevisiae*, which were 65% for a sextuple gene deletion (Mans et al., 2015) and 96% for a quadruple deletion (Ferreira et al., 2018). Recently, a CRISPR-Cpf1 system was shown to be very efficient for simultaneous gene disruption in *S. cerevisiae*: efficiency levels reached 88% and 100% for a quadruple deletion (genomic integration of Cpf1 vs. Cpf1 expression from a multicopy plasmid, respectively; Swiat et al., 2017). With further development, these strategies could serve as helpful tools for increasing the efficacy of multitarget mutations in *Y. lipolytica*.

5. Host strains

Y. lipolytica displays interstrain variability: different strains have different morphologies and metabolite patterns as they grow (Egermeler et al., 2017); they may also have genomic differences (Naumova et al., 1993). As previously underscored in the literature, in metabolic engineering, the choice of the parental strain is key to optimum system performance (Abghari et al., 2017; Larroude et al., 2017; Steensels et al., 2014; Xie et al., 2015). A summary of the most commonly used strains is presented below and in Table 2.

Table 2

Most commonly used *Yarrowia lipolytica* strains.

Strain	Genotype and characteristics	References
E150	*MATb his-1 leu2–270, ura3–302, xpr2–322, pXPR2-SUC2* reference strain, genome sequenced, grown on sucrose, Leu−, Ura−	Barth and Gaillardin, 1996
W29	*MATa*, French wild-type strain sequenced genome	
H222	*MATa*, German wild-type strain	
CBS6124	*MATa*, American wild-type strain	
Po1d	*MATa, leu2–270, ura3–302, xpr2–322, pXPR2-SUC2* derived from W29, extracellular protease AEP deleted, grown on sucrose, Leu−, Ura−	Le Dall et al., 1994
Po1f	*MATa, leu2–270, ura3–302, xpr2–322, axp1–2, pXPR2-SUC2* derived from Po1d, both extracellular proteases deleted, grown on sucrose, Leu−, Ura−	Madzak et al., 2000
Po1g	*MATa, leu2–270, ura3–302::URA3, xpr2–322, axp1–2, pXPR2-SUC2* derived from Po1f, both extracellular proteases deleted, grown on sucrose, pBR docking platform, Leu−	Madzak et al., 2000
Po1h	*MATa, ura3–302, xpr2–322, axp1–2, pXPR2-SUC2* derived from Po1f, both extracellular proteases deleted, grown on sucrose, Ura−	Madzak, 2003
Y1212	*MATa, leu2–270, ura3–302, xpr2–322, lip2Δ, lip7Δ, lip8Δ, Leu2-Zeta, pXPR2-SUC2* derived from Po1d, zeta platform, lipΔ, the three main lipases deleted, Ura−	Bordes et al., 2007

The most widely used *Y. lipolytica* genetic backgrounds are the wild-type French strain W29 (CLIB89), the wild-type German strain H222 (DSM 27185), the wild-type American strain CBS6124–2, and the wild-type Polish strain A101 (Barth and Gaillardin, 1996; Wojtatowicz et al., 1991). The reference strain for *Y. lipolytica* is E150 (CLIB122), whose genome has been fully sequenced and annotated (Dujon et al., 2004). This strain was derived from multiple back-crosses between the W29 and CBS6124–2 strains (Barth and Gaillardin, 1996). The genome sequence of W29 is now also available (Magnan et al., 2016). Genetic and physiological differences are naturally present among the strains. First, in contrast to the European strains, the American strain has a retrotransposon dispersed throughout its genome, Ylt1, which can also occur as solo zeta elements (Ylt1's long terminal repeats) (Barth and Gaillardin, 1996; Schmid-Berger et al., 1994). As mentioned above, zeta elements function as HR sites, and, due to their high copy number in the genome, they can be used as a target for the multiple integration of exogenous DNA (Juretzek et al., 2001; Pignede et al., 2000). Second, H222 naturally overproduces α-ketoglutarate, which is used in industrial processes as a building block for synthesizing heterocycles and also serves as a dietary supplement (Yovkova et al., 2014). Additionally, strains can differ in their filamentation profiles (Barth and Gaillardin, 1997; Thevenieau et al., 2009). The hyphal form can be 100% undesirable, causing fermenter clogs or low production yields, or 100% necessary, such as in immobilized cell bioreactors (Vandermies et al., 2018).

The most commonly used recipient strains are Po1 series strains (Po1d, f, g, and h), which were derived from the wild-type strain W29 (Le Dall et al., 1994; Madzak et al., 2000). They have been engineered to express the heterologous gene *SUC2* from *S. cerevisiae*, which allows sucrose to be used as a carbon source. This trait is of particular interest in industrial applications because it means yeast can exploit molasses, a cheap and abundant agroindustrial substrate (Nicaud et al., 1989a, 1989b). Both extracellular proteases (AEP and AXP) have been deleted from Po1f, g, and h, making them more suitable for heterologous protein expression (Madzak et al., 2000). Furthermore, Po1d and H222 derivatives, in which *Ku70* and/or *Ku80* were deleted, were constructed to increase HR efficiency (Kretzschmar et al., 2013; Verbeke et al., 2013).

Some derivative strains have also been built for more specific applications. The Y1212 strain, derived from Po1d, is equipped with an integrated zeta docking platform, for facilitating the incorporation of zeta-based integrative vectors. The three main lipase-encoding genes (*LIP2, LIP7* and *LIP8*) have been deleted in this strain, which makes it suitable for genetically engineering lipid metabolic pathways (Bordes et al., 2007). This strain was developed to be able to efficiently compare activity levels among enzymes or variants. Activity can be quantified directly in the supernatant, rendering protein purification and quantification unnecessary. This is possible because of expression level reproducibility, attributable to the integration of a single copy of the expression cassette at the zeta docking platform. Therefore, activity patterns are due to enzymatic differences at the molecular level and not to differences in expression levels (Cambon et al., 2010).

Other *Y. lipolytica* chassy strains of interest include strains for producing humanized glycoproteins, which help bypass the immunogenic problems associated with recombinant therapeutic proteins (De Pourcq et al., 2012a, 2012b) or strains for bioconverting cellulose, an abundant and renewable carbon source, into products of commercial interest (Guo et al., 2018). Strains have also been engineered to take advantage of a greater substrate range, allowing cheaper substrates to be used as carbon sources with a view to reducing fermentation costs. These strains were recently reviewed by Ledesma-Amaro and Nicaud (2016a).

It is clear that *Y. lipolytica* strains display a broad range of features, and several options are available when selecting the appropriate host strain. However, other strains of interest remain to be developed, such as strains with a disrupted *EYK1* gene, which would allow the use of the inducible promotor pEYK (described above), or strains resistant to high concentrations of glucose and glycerol.

It would also be interesting to better characterize the physiological features of the *Y. lipolytica* wild-types and to evaluate their natural traits as they relate to such biological processes as filamentation, intermediate metabolite production (e.g., of citric acid, succinate, alpha-ketoglutarate, and erythritol), lipid production, and resistance to antibiotics. As shown by Egermeier et al. (2017), environmental conditions can also have an impact. The researchers examined the ability of different strains to convert glycerol into polyols and citric acid under two different pH conditions; significant differences in metabolite patterns were observed. This finding shows that metabolism is dependent on the environment, a discovery that was also reported by Tomaszewska et al. (2014, 2012). This knowledge helps clarify *Y. lipolytica*'s physiology and the underlying regulatory mechanisms, information that will inform strain choice, which will depend on the desired application. As new synthetic biology techniques, genome-editing tools, and expression cassette insertion methods become available, it will become quicker and easier to transfer metabolically engineered modifications into other wild-type strains.

6. Computational tools

The advent of whole-genome sequencing and the reconstruction of metabolic networks at the genome scale have enabled the development of computer-assisted design tools that can be used to guide metabolic engineering. Such tools are important for building novel biosynthetic pathways and improving fluxes in existing pathways. More specifically, these models can be used to predict the outcomes of genetic

Table 3
Comparison of the different GEMs available for *Y. lipolytica*. The number of genes, reactions, and compartments were determined using the suite package sybil (Gelius-Dietrich et al., 2013). Accuracy values correspond to the values published by the authors. ND: not determined.

Name	No. genes	No. reactions	No. compartments	Accuracy	References
iNL895	898	1989	16	0.65	Loira et al., 2012
iYL619_PCP	619	1142	2	0.83	Pan and Hua, 2012
iMK735	735	1337	8	0.8	Kavšček et al., 2015
iYali4	901	1985	16	ND	Kerkhoven et al., 2016
iYLI647	646	1343	8	ND	Mishra et al., 2018

modifications and metabolic responses to environmental conditions as well as to determine optimal engineering strategies (Fernández-Castané et al., 2014; Oberhardt et al., 2009). The construction of accurate genome-scale metabolic models (GEMs) is crucial for properly simulating cell behavior, which requires high-quality annotated genome sequences and experimental data (Aung et al., 2013). Five GEMs have been developed thus far for *Y. lipolytica* (Table 3), which we will describe below.

Y. lipolytica was fully sequenced in 2004, as part of the Génolevures program, and the quality of manual annotation is high (Dujon et al., 2004; Sherman et al., 2009). The first functional GEMs were built by Loria and colleagues in 2012 (Loira et al., 2012), by combining in silico tools and manual curation; using a *S. cerevisiae* model as a scaffold; and validating efforts using previously published experimental data. For the *Y. lipolytica* iNL895 model, there was a fair degree of concordance between the model's growth predictions and the experimental results (accuracy: 0.65).

At almost the same time, another GEM was developed by Pan and Hua (Pan and Hua, 2012). In this case, the metabolic network, iY619_PCP, was reconstructed using genome annotation and information from biochemical databases such as KEGG, ENZYME, and BIGG. The in silico model successfully predicted growth in minimal media and on different substrates (accuracy: 0.83). The authors also used flux balance analysis (FBA) with single-gene knockouts to predict gene essentiality. Flux variability analysis (FVA) was employed to design new mutant strains that redirect fluxes toward lipid production.

Another GEM model for *Y. lipolytica*, named iMK735, was created based on a *S. cerevisiae* model (iND750) by Kavšček et al. (2015). The model was manually curated for species-specific reactions, and then FBA was used to design fermentation strategies where lipid production was optimized. Concordance between model predictions and experimental results was high (accuracy: 0.80). In addition, the model correctly predicted that a reduced aeration rate would induce lipid accumulation.

A fourth GEM model, iYali4 (Kerkhoven et al., 2016), was constructed using the recently described Yeast 7.11 consensus network (Aung et al., 2013) and curated to include unique reactions from both iYL619_PCP and iNL895. It was used to study the regulation of lipid metabolism. The researchers carried out integrative analysis of multi-level omics data obtained from *Y. lipolytica* chemostat cultures grown under carbon- and nitrogen-limited conditions. They showed that the previously documented increase in lipid accumulation after nitrogen depletion was not regulated at the transcriptional level but, instead, was related to amino-acid metabolism.

Very recently, Mishra and colleagues built a new objective-oriented model for simulating long-chain dicarboxylic acid (DCA) production. This new GEM, named iYLI647, was constructed using the iMK735 model as a scaffold, and manual curation was performed to expand the model's characteristics. For example, reactions from the ω–oxidation and the β –oxidation pathways were incorporated; the model also allowed the separation of biomass synthesis equations for growth under carbon- and nitrogen-limited conditions, enhancing the accuracy of growth predictions relative to previous models. The model was then used to identify genetic engineering targets with DCA overproduction in mind (Mishra et al., 2018).

Traditional GEMs have limitations as they incorporate only stoichiometric constraints; they assume steady-state metabolite concentrations. Thus, efforts are being made to integrate kinetic information into the models. Several dynamic mathematical models have been developed to help optimize lipid contents and/or to study citric acid production in batch and continuous processes (Arzumanov et al., 2000; Papanikolaou and Aggelis, 2003a, 2003b; Papanikolaou et al., 2006). However, these models still do not consider internal regulation of metabolism and metabolic shifts.

Recently, Robles-Rodriguez et al. (2017) came up with three dynamic metabolic models, based on a simplified metabolic network, for describing lipid accumulation and citric acid production by *Y. lipolytica* growing on glucose. These models can guide the design of strategies for improving culture performance via the identification of rate-controlling steps and metabolic fluxes. Three independent experimental data sets, obtained from fed-batch and sequential-batch cultures of *Y. lipolytica* grown on glucose under conditions of nitrogen limitation and deficiency, were used to calibrate and validate the models. The models' predictions were reasonably close to the experimental data. A common advantage of these dynamic metabolic models is that they can incorporate metabolic descriptions and regulation mechanisms and can thus be used to identify rate-controlling steps in the metabolic network.

At a higher scale, high-throughput technologies make it possible to analyze large amounts of omics data, facilitating the investigation of cell metabolism and physiology at the systems level. In recent years, such research has been carried out in *Y. lipolytica*. For instance, transcriptome analysis revealed the existence of four different transcription profiles over a 32-h fermentation period and identified genes potentially involved in the metabolism of oleaginous species (Morin et al., 2011). A separate transcriptome analysis, carried out in tandem with proteomics methods, was used to explore amino acid catabolism (Mansour et al., 2009; Morín et al., 2007). Additionally, proteome analyses were conducted in *Y. lipolytica* to characterize the proteins involved in the yeast-to-hypha transition (Morín et al., 2007); the osmotic response to erythritol (Yang et al., 2015); and the degradation of TNT (Khilyas et al., 2017). Fluxomics has grown as a discipline thanks to 13C-based metabolic flux analysis. The latter was used to discover that the pentose phosphate pathway is the major source of the cofactor required for lipid production (Wasylenko et al., 2015). The response to nitrogen limitation, and its effects on lipid storage regulation, was analyzed by Pomraning et al. (2016) using a multiomics approach. Another multiomics study found that carbon fluxes were redirected from amino acids to lipids in *Y. lipolytica* grown in carbon- and nitrogen-limited chemostat cultures (Kerkhoven et al., 2016). In subsequent work, the same researchers showed that leucine biosynthesis was particularly downregulated when the yeast was grown under nitrogen-limited conditions, concomitantly with lipid accumulation (Kerkhoven et al., 2017). A recent study looked at transcriptional changes in *Y. lipolytica* during lipid biosynthesis in strain carrying a *MHY1* gene inactivation (Wang et al., 2018); *MHY1* encodes a C_2H_2-type zinc finger protein. They found that nearly 25% of annotated *Y. lipolytica* genes were expressed at significantly different levels, suggesting Mhyp plays a crucial regulatory role in various biological processes, including lipid and amino acid metabolism (again underscoring the interaction between these pathways). Trebulle and colleagues

(2017) used transcriptomic data gathered during lipid accumulation to infer the gene regulatory network; the goal was to identify regulators involved in lipid accumulation. The nine highest ranked transcription factors were then overexpressed in a wild-type strain over the course of a systematic high-throughput functional analysis carried out by Leplat et al. (2018); overall, 148 putative transcription factors were overexpressed in this study. For six of the nine mutants obtained, lipid content was at least 10% greater than that in the wild-type, which validates the utility of the GRN approach for identifying context-specific transcription factors.

Genome-scale modeling contributes significantly to our understanding of cellular processes and is very useful for guiding metabolic engineering via the improvement of strain performance. However, since there are differences in GEM curation procedures and coverage of metabolites and reactions differs, the models vary in prediction accuracy. The optimal model will depend on one's objectives. That said, efforts should continue to improve the models and their predictions. For example, additional components, such as omics data and/or enzyme kinetics and abundance, could be included in GEMs, as it has been recently done for *S. cerevisiae* (Sánchez et al., 2017).

7. Conclusions and perspectives

In this review, we have attempted to describe the state-of-the-art synthetic biology tools available for *Y. lipolytica*, a micro-organism of industrial importance. We also discussed the most commonly used DNA parts and genetic engineering techniques used with this yeast. Some of these methods have the distinct potential to become standard lab techniques for *Y. lipolytica*. The greatest promise is held by tools and methods for identifying and characterizing new parts, carrying out DNA assembly, and performing genome editing.

It is important to note that many more synthetic biology tools are available for the model organisms *S. cerevisiae* and *E. coli* because they are well characterized, grow quickly, and are easy to stably transform. However, other yeasts are often more desirable as bioprocessing hosts because their natural metabolisms render them more suitable for producing the target product. With the development of new synthetic biology tools, some previously non-conventional yeasts, such as *Y. lipolytica* or *Pichia pastoris*, are on their way to becoming model organisms (Gellissen et al., 2005; Löbs et al., 2017; Wagner and Alper, 2016). The impact of these tools is evidenced by the increase in the number of engineered strains. Other yeasts, like *Kluyveromyces marxianus* and *Rhodosporidium toruloides*, have certain features that make them suitable for industrial applications. However, their molecular tools are less well developed. This state is expected to change in the coming years as the development of synthetic biology tools continues (Lane and Morrissey, 2010; Park et al., 2018a, 2018b).

Y. lipolytica is considered to be non-pathogenic and is generally regarded as safe. Its ability to generate large amounts of biomass on simple substrates makes it a good host for producing pharmaceutical compounds and food additives, among other products, especially given the availability of its fully sequenced genome and diverse metabolic engineering toolkit. The synthetic biology tools described in this review do not only serve as proof of concept; in many cases, they have already been used to engineer *Y. lipolytica* strains. Several recent reviews discuss the products and production yields obtained with these strains (Darvishi et al., 2018; Shabbir Hussain et al., 2016; Ledesma-Amaro and Nicaud, 2016b; Liu et al., 2015; Madzak, 2018, 2015; Markham and Alper, 2018; Xie, 2017), underscoring *Y. lipolytica*'s potential as a cell factory.

Synthetic biology is an emerging engineering discipline and, as such, applies key engineering concepts such as standardization, modularity, predictability, reliability, and modeling in its work with biological systems (Andrianantoandro et al., 2006).

Most of the standard promoters and other DNA parts used for engineering *Y. lipolytica* have not been studied rigorously enough to meet synthetic biology standards for predictability and reliability. For example, parts must be characterized under standardized conditions, and strain background must be taken into consideration. While efficient wild-type and hybrid promoters have been developed, such as the strong constitutive pTEF, the phase-dependent hp4d, the fatty-acid-inducible pPOX2, and the erythritol/erythrulose-inducible promoter pEYK1, more inducible promoters are needed so that gene expression can be switched on and off. In addition, it is crucial to confirm the reproducibility of the promoter-related results obtained by different labs.

Modularity in *Y. lipolytica* is also currently being explored (although it is still in an early stage of development), thanks to the recent creation of modular cloning systems such as the above-mentioned Golden Gate system (Celińska et al., 2017). It would be highly beneficial to add novel, fully characterized parts to this toolbox.

Models of *Y. lipolytica* metabolism keep increasing in number, and it is difficult for non-experts to evaluate which model is the most appropriate for a given application. This challenge has been partially dealt with in other organisms such as *S. cerevisiae*, where researchers joined together to create a consensus model (Herrgård et al., 2008). Such a strategy could be applied for *Y. lipolytica*: taking the best parts of existing models to create a baseline model that could be improved through refinement.

Although many additional advances are needed before *Y. lipolytica* can obtain the status of a model organism, significant efforts are being made so that this yeast can be fully exploited. Consequently, it is expected that new research will continue to improve the techniques describe herein and to develop innovative tools and technologies aimed at better engineering *Y. lipolytica* strains.

Acknowledgments

References

Abghari, A., Chen, S., Abghari, A., Chen, S., 2017. Engineering Yarrowia lipolytica for Enhanced production of Lipid and Citric Acid. Fermentation 3, 34. https://doi.org/10.3390/fermentation3030034.

Andrianantoandro, E., Basu, S., Karig, D.K., Weiss, R., 2006. Synthetic biology: new engineering rules for an emerging discipline. Mol. Syst. Biol. 2, 2006.0028. https://doi.org/10.1038/msb4100073.

Arzumanov, T.E., Sidorov, I.A., Shishkanova, N.V., Finogenova, T.V., 2000. Mathematical modeling of citric acid production by repeated batch culture. Enzym. Microb. Technol. 26, 826–833.

Aung, H.W., Henry, S.A., Walker, L.P., 2013. Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of yeast Metabolism. Ind. Biotechnol. 9, 215–228.

Bakkaiova, J., Arata, K., Matsunobu, M., Ono, B., Aoki, T., Lajdova, D., Nebohacova, M., Nosek, J., Miyakawa, I., Tomaska, L., 2014. The strictly Aerobic yeast Yarrowia lipolytica Tolerates loss of a Mitochondrial DNA-Packaging Protein. Eukaryot. Cell 13, 1143–1157. https://doi.org/10.1128/EC.00092-14.

Barth, G., Gaillardin, C., 1996. Yarrowia lipolytica. In: Nonconventional Yeasts in Biotechnology, pp. 313–388.

Barth, G., Gaillardin, C., 1997. Physiology and genetics of the dimorphic fungus Yarrowia lipolytica. FEMS Microbiol. Rev. 19, 219–237.

Beopoulos, A., Verbeke, J., Bordes, F., Guicherd, M., Bressy, M., Marty, A., Nicaud, J.-M., 2014. Metabolic engineering for ricinoleic acid production in the oleaginous yeast Yarrowia lipolytica. Appl. Microbiol. Biotechnol. 98, 251–262. https://doi.org/10.1007/s00253-013-5295-x.

Bhutada, G., Kavšček, M., Ledesma-Amaro, R., Thomas, S., Rechberger, G.N., Nicaud, J.-M., Natter, K., 2017. Sugar versus fat: elimination of glycogen storage improves lipid accumulation in Yarrowia lipolytica. FEMS Yeast Res. 17. https://doi.org/10.1093/femsyr/fox020.

Blazeck, J., Liu, L., Redden, H., Alper, H., 2011. Tuning gene expression in Yarrowia lipolytica by a hybrid promoter approach. Appl. Environ. Microbiol. 77, 7905–7914.

Blazeck, J., Reed, B., Garg, R., Gerstner, R., Pan, A., Agarwala, V., Alper, H.S., 2013. Generalizing a hybrid synthetic promoter approach in Yarrowia lipolytica. Appl.

Microbiol. Biotechnol. 97, 3037–3052.

Blazeck, J., Hill, A., Jamoussi, M., Pan, A., Miller, J., Alper, H.S., 2015. Metabolic engineering of Yarrowia lipolytica for itaconic acid production. Metab. Eng. 32, 66–73.

Bordes, F., Fudalej, F., Dossat, V., Nicaud, J.-M., Marty, A., 2007. A new recombinant protein expression system for high-throughput screening in the yeast Yarrowia lipolytica. J. Microbiol. Methods 70, 493–502.

Braga, A., Belo, I., 2016. Biotechnological production of γ-decalactone, a peach like aroma, by Yarrowia lipolytica. World J. Microbiol. Biotechnol. 32, 169.

Bredeweg, E.L., Pomraning, K.R., Dai, Z., Nielsen, J., Kerkhoven, E.J., Baker, S.E., 2017. A molecular genetic toolbox for Yarrowia lipolytica. Biotechnol. Biofuels 10, 2.

Bulani, S.I., Moleleki, L., Albertyn, J., Moleleki, N., 2012. Development of a novel rDNA based plasmid for enhanced cell surface display on Yarrowia lipolytica. AMB Express 2, 27.

Cambon, E., Piamtongkam, R., Bordes, F., Duquesne, S., Laguerre, S., Nicaud, J.-M., Marty, A., 2010. A new Yarrowia lipolytica expression system: an efficient tool for rapid and reliable kinetic analysis of improved enzymes. Enzym. Microb. Technol. 47, 91–96. https://doi.org/10.1016/j.enzmictec.2010.06.001.

Campos-Góngora, E., Andaluz, E., Bellido, A., Ruiz-Herrera, J., Larriba, G., 2013. The RAD52 ortholog of Yarrowia lipolytica is essential for nuclear integrity and DNA repair. FEMS Yeast Res. 13, 441–452. https://doi.org/10.1111/1567-1364.12047.

Carly, F., Vandermies, M., Telek, S., Steels, S., Thomas, S., Nicaud, J.-M., Fickers, P., 2017. Enhancing erythritol productivity in Yarrowia lipolytica using metabolic engineering. Metab. Eng. 42, 19–24. https://doi.org/10.1016/j.ymben.2017.05.002.

Casaregola, S., Feynerol, C., Diez, M., Fournier, P., Gaillardin, C., 1997. Genomic Organization of the Yeast Yarrowia lipolyticaChromosoma. Vol. 106. pp. 380–390.

Cavallo, E., Charreau, H., Cerrutti, P., Foresti, M.L., 2017. Yarrowia lipolytica: a model yeast for citric acid production. FEMS Yeast Res. 17. https://doi.org/10.1093/femsyr/fox084.

Celińska, E., Grajek, W., 2013. A novel multigene expression construct for modification of glycerol metabolism in Yarrowia lipolytica. Microb. Cell Factories 12, 102.

Celińska, E., Kubiak, P., Białas, W., Dziadas, M., Grajek, W., 2013. Yarrowia lipolytica: the novel and promising 2-phenylethanol producer. J. Ind. Microbiol. Biotechnol. 40, 389–392. https://doi.org/10.1007/s10295-013-1240-3.

Celińska, E., Ledesma-Amaro, R., Larroude, M., Rossignol, T., Pauthenier, C., Nicaud, J.-M., 2017. Golden Gate Assembly system dedicated to complex pathway manipulation in Yarrowia lipolytica. Microb. Biotechnol. 10, 450–455.

Celińska, E., Borkowska, M., Białas, W., Korpys, P., Nicaud, J., 2018. Robust signal peptides for protein secretion in Yarrowia lipolytica: identification and characterization of novel secretory tags. Appl. Microbiol. Biotechnol. 102, 5221–5233.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J., Voytas, D.F., 2010. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186, 757–761.

Cui, W., Wang, Q., Zhang, F., Zhang, S.-C., Chi, Z.-M., Madzak, C., 2011. Direct conversion of inulin into single cell protein by the engineered Yarrowia lipolytica carrying inulinase gene. Process Biochem. 46, 1442–1448.

Curran, K.A., Morse, N.J., Markham, K.A., Wagman, A.M., Gupta, A., Alper, H.S., 2015. Short Synthetic Terminators for improved Heterologous Gene Expression in yeast. ACS Synth. Biol. 4, 824–832.

Dall, M.-T.L., Nicaud, J.-M., Gaillardin, C., 1994. Multiple-copy integration in the yeast Yarrowia lipolytica. Curr. Genet. 26, 38–44. https://doi.org/10.1007/BF00326302.

Darvishi, F., Fathi, Z., Ariana, M., Moradi, H., 2017. Yarrowia lipolytica as a workhorse for biofuel production. Biochem. Eng. J. 127, 87–96.

Darvishi, F., Ariana, M., Marella, E.R., Borodina, I., 2018. Advances in synthetic biology of oleaginous yeast Yarrowia lipolytica for producing non-native chemicals. Appl. Microbiol. Biotechnol. 102, 5925–5938. https://doi.org/10.1007/s00253-018-9099-x.

De Pourcq, K., Tiels, P., Van Hecke, A., Geysens, S., Vervecken, W., Callewaert, N., 2012a. Engineering Yarrowia lipolytica to produce glycoproteins homogeneously modified with the universal Man3GlcNAc2 N-glycan core. PLoS One 7, e39976. https://doi.org/10.1371/journal.pone.0039976.

De Pourcq, K., Vervecken, W., Dewerte, I., Valevska, A., Van Hecke, A., Callewaert, N., 2012b. Engineering the yeast Yarrowia lipolytica for the production of therapeutic proteins homogeneously glycosylated with Man8GlcNAc2 and Man5GlcNAc2. Microb. Cell Factories 11, 53. https://doi.org/10.1186/1475-2859-11-53.

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuvéglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.-M., Beyne, E., Bleykasten, C., Boisramé, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.-M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard, G.-F., Straub, M.-L., Suleau, A., Swennen, D., Tekaia, F., Wésolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P., Souciet, J.-L., 2004. Genome evolution in yeasts. Nature 430, 35–44.

Dulermo, R., Brunel, F., Dulermo, T., Ledesma-Amaro, R., Vion, J., Trassaert, M., Thomas, S., Nicaud, J.-M., Leplat, C., 2017. Using a vector pool containing variable-strength promoters to optimize protein production in Yarrowia lipolytica. Microb. Cell Factories 16, 31.

Duquesne, S., Bozonnet, S., Bordes, F., Dumon, C., Nicaud, J.-M., Marty, A., 2014. Construction of a Highly active Xylanase Displaying Oleaginous yeast: Comparison of Anchoring Systems. PLoS One 9, e95128. https://doi.org/10.1371/journal.pone.0095128.

Egermeier, M., Russmayer, H., Sauer, M., Marx, H., 2017. Metabolic Flexibility of Yarrowia lipolytica growing on Glycerol. Front. Microbiol. 8. https://doi.org/10.3389/fmicb.2017.00049.

Engler, C., Kandzia, R., Marillonnet, S., 2008. A one pot, one step, precision cloning method with high throughput capability. PLoS One 3, e3647.

Fernández-Castané, A., Fehér, T., Carbonell, P., Pauthenier, C., Faulon, J.-L., 2014. Computer-aided design for metabolic engineering. J. Biotechnol. 192, 302–313.

Ferreira, R., Skrekas, C., Nielsen, J., David, F., 2018. Multiplexed CRISPR/Cas9 Genome Editing and Gene Regulation using Csy4 in Saccharomyces cerevisiae. ACS Synth. Biol. 7, 10–15. https://doi.org/10.1021/acssynbio.7b00259.

Fickers, P., Le Dall, M.T., Gaillardin, C., Thonart, P., Nicaud, J.M., 2003. New disruption cassettes for rapid gene disruption and marker rescue in the yeast Yarrowia lipolytica. J. Microbiol. Methods 55, 727–737.

Fournier, P., Abbas, A., Chasles, M., Kudla, B., Ogrydziak, D.M., Yaver, D., Xuan, J.W., Peito, A., Ribet, A.M., Feynerol, C., 1993. Colocalization of centromeric and replicative functions on autonomously replicating sequences isolated from the yeast Yarrowia lipolytica. Proc. Natl. Acad. Sci. U. S. A. 90, 4912–4916.

Franke, A.E., Kaczmarek, F.S., Eisenhard, M.E., Geoghegan, K.F., Danley, D.E., De Zeeuw, J.R., O'Donnell, M.M., Gollaher, M.G.J., Davidow, L.S., 1988. Expression and Secretion of Bovine Prochymosin in Yarrowia Lipolytica. (Developments in industrial microbiology (USA).

Friedlander, J., Tsakraklides, V., Kamineni, A., Greenhagen, E.H., Consiglio, A.L., MacEwen, K., Crabtree, D.V., Afshar, J., Nugent, R.L., Hamilton, M.A., Joe Shaw, A., South, C.R., Stephanopoulos, G., Brevnova, E.E., 2016. Engineering of a high lipid producing Yarrowia lipolytica strain. Biotechnology for Biofuels 9 (77). https://doi.org/10.1186/s13068-016-0492-3.

Gaj, T., Gersbach, C.A., Barbas 3rd, C.F., 2013. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31, 397–405.

Gallegos, J.E., Rose, A.B., 2015. The enduring mystery of intron-mediated enhancement. Plant Sci. Int. J. Exp. Plant Biol. 237, 8–15. https://doi.org/10.1016/j.plantsci.2015.04.017.

Gao, S., Han, L., Zhu, L., Ge, M., Yang, S., Jiang, Y., Chen, D., 2014. One-step integration of multiple genes into the oleaginous yeast Yarrowia lipolytica. Biotechnol. Lett. 36, 2523–2528.

Gao, C., Qi, Q., Madzak, C., Lin, C.S.K., 2015. Exploring medium-chain-length polyhydroxyalkanoates production in the engineered yeast Yarrowia lipolytica. J. Ind. Microbiol. Biotechnol. 42, 1255–1262. https://doi.org/10.1007/s10295-015-1649-y.

Gao, S., Tong, Y., Wen, Z., Zhu, L., Ge, M., Chen, D., Jiang, Y., Yang, S., 2016. Multiplex gene editing of the Yarrowia lipolytica genome using the CRISPR-Cas9 system. J. Ind. Microbiol. Biotechnol. 43, 1085–1093.

Gao, D., Smith, S., Spagnuolo, M., Rodriguez, G., Blenner, M., 2018. Dual CRISPR-Cas9 Cleavage mediated Gene excision and Targeted Integration in Yarrowia lipolytica. Biotechnol J e1700590. https://doi.org/10.1002/biot.201700590.

Geisberg, J.V., Moqtaderi, Z., Fan, X., Ozsolak, F., Struhl, K., 2014. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. Cell 156, 812–824.

Gelius-Dietrich, G., Desouki, A.A., Fritzemeier, C.J., Lercher, M.J., 2013. Sybil – Efficient constraint-based modelling in R. BMC Syst. Biol. 7, 125. https://doi.org/10.1186/1752-0509-7-125.

Gellissen, G., Kunze, G., Gaillardin, C., Cregg, J.M., Berardi, E., Veenhuis, M., van der Klei, I., 2005. New yeast expression platforms based on methylotrophic Hansenula polymorpha and Pichia pastoris and on dimorphic Arxula adeninivorans and Yarrowia lipolytica– a comparison. FEMS Yeast Res. 5, 1079–1096. https://doi.org/10.1016/j.femsyr.2005.06.004.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Iii, C.A.H., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345. https://doi.org/10.1038/nmeth.1318.

Guo, Z., Duquesne, S., Bozonnet, S., Cioci, G., Nicaud, J.-M., Marty, A., O'Donohue, M.J., 2015. Development of cellobiose-degrading ability in Yarrowia lipolytica strain by overexpression of endogenous genes. Biotechnol. Biofuels 8 (109). https://doi.org/10.1186/s13068-015-0289-9.

Guo, H., Su, S., Madzak, C., Zhou, J., Chen, H., Chen, G., 2016. Applying pathway engineering to enhance production of alpha-ketoglutarate in Yarrowia lipolytica. Appl. Microbiol. Biotechnol. 100, 9875–9884.

Guo, Z., Robin, J., Duquesne, S., O'Donohue, M.J., Marty, A., Bordes, F., 2018. Developing cellulolytic Yarrowia lipolytica as a platform for the production of valuable products in consolidated bioprocessing of cellulose. Biotechnol. Biofuels 11 (141). https://doi.org/10.1186/s13068-018-1144-6.

Haddouche, R., Delessert, S., Sabirova, J., Neuvéglise, C., Poirier, Y., Nicaud, J.-M., 2010. Roles of multiple acyl-CoA oxidases in the routing of carbon flow towards β-oxidation and polyhydroxyalkanoate biosynthesis in Yarrowia lipolytica. FEMS Yeast Res. 10, 917–927.

Haddouche, R., Poirier, Y., Delessert, S., Sabirova, J., Pagot, Y., Neuvéglise, C., Nicaud, J.-M., 2011. Engineering polyhydroxyalkanoate content and monomer composition in the oleaginous yeast Yarrowia lipolytica by modifying the ß-oxidation multifunctional protein. Appl. Microbiol. Biotechnol. 91, 1327–1340. https://doi.org/10.1007/s00253-011-3331-2.

Han, Z., Madzak, C., Su, W.W., 2013. Tunable nano-oleosomes derived from engineered Yarrowia lipolytica. Biotechnol. Bioeng. 110, 702–710.

Hanko, E.K.R., Denby, C.M., Sànchez, I., Nogué, V., Lin, W., Ramirez, K.J., Singer, C.A., Beckham, G.T., Keasling, J.D., 2018. Engineering β-oxidation in Yarrowia lipolytica for methyl ketone production. Metab. Eng. 48, 52–62. https://doi.org/10.1016/j.ymben.2018.05.018.

Hartley, J.L., 2000. DNA Cloning using in Vitro Site-specific Recombination. Genome Res. 10, 1788–1795.

Herrgård, M.J., Swainston, N., Dobson, P., Dunn, W.B., Arga, K.Y., Arvas, M., Blüthgen, N., Borger, S., Costenoble, R., Heinemann, M., Hucka, M., Le Novère, N., Li, P., Liebermeister, W., Mo, M.L., Oliveira, A.P., Petranovic, D., Pettifer, S., Simeonidis, E.,

Smallbone, K., Spasić, I., Weichart, D., Brent, R., Broomhead, D.S., Westerhoff, H.V., Kirdar, B., Penttilä, M., Klipp, E., Palsson, B.Ø., Sauer, U., Oliver, S.G., Mendes, P., Nielsen, J., Kell, D.B., 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat. Biotechnol. 26 (10), 1155–1160. https://doi.org/10.1038/nbt1492.

Holkenbrink, C., Dam, M.I., Kildegaard, K.R., Beder, J., Dahlin, J., Belda, D.D., Borodina, I., 2018. EasyCloneYALI: CRISPR/Cas9-based Synthetic Toolbox for Engineering of the yeast Yarrowia lipolytica. Biotechnol. J. 13 (9), e1700543. https://doi.org/10.1002/biot.201700543. Epub 2018 Feb 14.

Hong, S.-P., Seip, J., Walters-Pollak, D., Rupert, R., Jackson, R., Xue, Z., Zhu, Q., 2012. Engineering Yarrowia lipolytica to express secretory invertase with strong FBA1IN promoter. Yeast 29, 59–72.

Imatoukene, N., Verbeke, J., Beopoulos, A., Idrissi Taghki, A., Thomasset, B., Sarde, C.-O., Nonus, M., Nicaud, J.-M., 2017. A metabolic engineering strategy for producing conjugated linoleic acids using the oleaginous yeast Yarrowia lipolytica. Appl. Microbiol. Biotechnol. 101, 4605–4616.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816–821.

Johnravindar, D., Karthikeyan, O.P., Selvam, A., Murugesan, K., Wong, J.W.C., 2018. Lipid accumulation potential of oleaginous yeasts: a comparative evaluation using food waste leachate as a substrate. Bioresour. Technol. 248, 221–228. https://doi.org/10.1016/j.biortech.2017.06.151.

Juretzek, T., Wang, H.-J., Nicaud, J.-M., Mauersberger, S., Barth, G., 2000. Comparison of promoters suitable for regulated overexpression of β-galactosidase in the alkane-utilizing yeast Yarrowia lipolytica. Biotechnol. Bioprocess Eng. 5, 320–326.

Juretzek, T., Le Dall, M., Mauersberger, S., Gaillardin, C., Barth, G., Nicaud, J., 2001. Vectors for gene expression and amplification in the yeast Yarrowia lipolytica. Yeast 18, 97–113.

Kavšček, M., Bhutada, G., Madl, T., Natter, K., 2015. Optimization of lipid production with a genome-scale model of Yarrowia lipolytica. BMC Syst. Biol. 9, 72.

Kerkhoven, E.J., Pomraning, K.R., Baker, S.E., Nielsen, J., 2016. Regulation of amino-acid metabolism controls flux to lipid accumulation in Yarrowia lipolytica. NPJ Syst Biol Appl 2, 16005.

Kerkhoven, E.J., Kim, Y.-M., Wei, S., Nicora, C.D., Fillmore, T.L., Purvine, S.O., Webb-Robertson, B.-J., Smith, R.D., Baker, S.E., Metz, T.O., Nielsen, J., 2017. Leucine Biosynthesis is involved in Regulating High Lipid Accumulation in Yarrowia lipolytica. MBio 8 (e00857–17).

Kerscher, S., Bénit, P., Abdrakhmanova, A., Zwicker, K., Rais, I., Karas, M., Rustin, P., Brandt, U., 2004. Processing of the 24 kDa subunit mitochondrial import signal is not required for assembly of functional complex I in Yarrowia lipolytica. Eur. J. Biochem. 271, 3588–3595. https://doi.org/10.1111/j.0014-2956.2004.04296.x.

Khilyas, I.V., Lochnit, G., Ilinskaya, O.N., 2017. Proteomic Analysis of 2,4,6-Trinitrotoluene Degrading yeast Yarrowia lipolytica. Front. Microbiol. 8. https://doi.org/10.3389/fmicb.2017.02600.

Kjaergaard, M., Nilsson, C., Secher, A., Kildegaard, J., Skovgaard, T., Nielsen, M.O., Grove, K., Raun, K., 2017. Differential hypothalamic leptin sensitivity in obese rat offspring exposed to maternal and postnatal intake of chocolate and soft drink. Nutr. Diabetes 7, e242.

Kretzschmar, A., Otto, C., Holz, M., Werner, S., Hübner, L., Barth, G., 2013. Increased homologous integration frequency in Yarrowia lipolytica strains defective in non-homologous end-joining. Curr. Genet. 59, 63–72.

Lane, M.M., Morrissey, J.P., 2010. Kluyveromyces marxianus: a yeast emerging from its sister's shadow. Fungal Biology Reviews 24, 17–26. https://doi.org/10.1016/j.fbr.2010.01.001.

Larroude, M., Celinska, E., Back, A., Thomas, S., Nicaud, J.-M., Ledesma-Amaro, R., 2017. A synthetic biology approach to transform Yarrowia lipolytica into a competitive biotechnological producer of β-carotene. Biotechnol. Bioeng. https://doi.org/10.1002/bit.26473.

Lazar, Z., Rossignol, T., Verbeke, J., Coq, A.-M.C.-L., Nicaud, J.-M., Robak, M., 2013. Optimized invertase expression and secretion cassette for improving Yarrowia lipolytica growth on sucrose for industrial applications. J. Ind. Microbiol. Biotechnol. 40, 1273–1283.

Ledesma-Amaro, R., Nicaud, J.-M., 2016a. Metabolic Engineering for Expanding the Substrate Range of Yarrowia lipolytica. Trends Biotechnol. 34, 798–809. https://doi.org/10.1016/j.tibtech.2016.04.010.

Ledesma-Amaro, R., Nicaud, J.-M., 2016b. Yarrowia lipolytica as a biotechnological chassis to produce usual and unusual fatty acids. Prog. Lipid Res. 61, 40–50. https://doi.org/10.1016/j.plipres.2015.12.001.

Ledesma-Amaro, R., Dulermo, T., Nicaud, J.M., 2015. Engineering Yarrowia lipolytica to produce biodiesel from raw starch. Biotechnol. Biofuels 8, 148.

Ledesma-Amaro, R., Dulermo, R., Niehus, X., Nicaud, J.-M., 2016a. Combining metabolic engineering and process optimization to improve production and secretion of fatty acids. Metab. Eng. 38, 38–46.

Ledesma-Amaro, R., Lazar, Z., Rakicka, M., Guo, Z., Fouchard, F., Coq, A.-M.C.-L., Nicaud, J.-M., 2016b. Metabolic engineering of Yarrowia lipolytica to produce chemicals and fuels from xylose. Metab. Eng. 38, 115–124.

Leplat, C., Nicaud, J.-M., Rossignol, T., 2015. High-throughput transformation method for Yarrowia lipolytica mutant library screening. FEMS Yeast Res. 15. https://doi.org/10.1093/femsyr/fov052.

Leplat, C., Nicaud, J.-M., Rossignol, T., 2018. Overexpression screen reveals transcription factors involved in lipid accumulation in Yarrowia lipolytica. FEMS Yeast Res. 18. https://doi.org/10.1093/femsyr/foy037.

Li, Z.-J., Qiao, K., Liu, N., Stephanopoulos, G., 2017. Engineering Yarrowia lipolytica for poly-3-hydroxybutyrate production. J. Ind. Microbiol. Biotechnol. 44, 605–612. https://doi.org/10.1007/s10295-016-1864-1.

Liu, L., Otoupal, P., Pan, A., Alper, H.S., 2014. Increasing expression level and copy number of a Yarrowia lipolytica plasmid through regulated centromere function. FEMS Yeast Res. 14, 1124–1127. https://doi.org/10.1111/1567-1364.12201.

Liu, H.-H., Ji, X.-J., Huang, H., 2015. Biotechnological applications of Yarrowia lipolytica: past, present and future. Biotechnol. Adv. 33, 1522–1546. https://doi.org/10.1016/j.biotechadv.2015.07.010.

Löbs, A.-K., Schwartz, C., Wheeldon, I., 2017. Genome and metabolic engineering in non-conventional yeasts: current advances and applications. Synthetic and Systems Biotechnology 2, 198–207. https://doi.org/10.1016/j.synbio.2017.08.002.

Loira, N., Dulermo, T., Nicaud, J.-M., Sherman, D.J., 2012. A genome-scale metabolic model of the lipid-accumulating yeast Yarrowia lipolytica. BMC Syst. Biol. 6, 35.

Madzak, C., 2003. Recent research developments in microbiology Research. In: Pandalai, S.G. (Ed.), Signpost Trivandrum Vol7, pp. 453–479.

Madzak, C., 2015. Yarrowia lipolytica: recent achievements in heterologous protein expression and pathway engineering. Appl. Microbiol. Biotechnol. 99, 4559–4577.

Madzak, C., 2018. Engineering Yarrowia lipolytica for use in Biotechnological applications: a Review of Major Achievements and recent Innovations. Mol. Biotechnol. 60, 621–635. https://doi.org/10.1007/s12033-018-0093-4.

Madzak, C., Beckerich, J.-M., 2013. Heterologous Protein Expression and Secretion in Yarrowia lipolytica. Microbiology Monographs. 1–76.

Madzak, C., Tréton, B., Blanchin-Roland, S., 2000. Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast Yarrowia lipolytica. J. Mol. Microbiol. Biotechnol. 2, 207–216.

Magdouli, S., Guedri, T., Tarek, R., Brar, S.K., Blais, J.F., 2017. Valorization of raw glycerol and crustacean waste into value added products by Yarrowia lipolytica. Bioresour. Technol. 243, 57–68. https://doi.org/10.1016/j.biortech.2017.06.074.

Magnan, C., Yu, J., Chang, I., Jahn, E., Kanomata, Y., Wu, J., Zeller, M., Oakes, M., Baldi, P., Sandmeyer, S., 2016. Sequence Assembly of Yarrowia lipolytica Strain W29/CLIB89 shows Transposable Element Diversity. PLoS One 11, e0162363. https://doi.org/10.1371/journal.pone.0162363.

Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G.A., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J.A., Daran, J.-M.G., 2015. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae. FEMS Yeast Res. 15. https://doi.org/10.1093/femsyr/fov004.

Mansour, S., Bailly, J., Delettre, J., Bonnarme, P., 2009. A proteomic and transcriptomic view of amino acids catabolism in the yeast Yarrowia lipolytica. Proteomics 9, 4714–4725.

Markham, K.A., Alper, H.S., 2018. Synthetic Biology Expands the Industrial potential of Yarrowia lipolytica. Trends Biotechnol. 0. https://doi.org/10.1016/j.tibtech.2018.05.004.

Markham, K.A., Palmer, C.M., Chwatko, M., Wagner, J.M., Murray, C., Vazquez, S., Swaminathan, A., Chakravarty, I., Lynd, N.A., Alper, H.S., 2018. Rewiring Yarrowia lipolytica toward triacetic acid lactone for materials generation. PNAS 201721203. https://doi.org/10.1073/pnas.1721203115.

Martinez-Vazquez, A., Gonzalez-Hernandez, A., Domínguez, A., Rachubinski, R., Riquelme, M., Cuellar-Mata, P., Guzman, J.C.T., 2013. Identification of the transcription factor Znc1p, which regulates the yeast-to-hypha transition in the dimorphic yeast Yarrowia lipolytica. PLoS One 8, e66790.

Matsuoka, M., Matsubara, M., Daidoh, H., Imanaka, T., Uchida, K., Aiba, S., 1993. Analysis of regions essential for the function of chromosomal replicator sequences from Yarrowia lipolytica. Mol. Gen. Genet. 237, 327–333.

Matthäus, F., Ketelhot, M., Gatter, M., Barth, G., 2014. Production of lycopene in the non-carotenoid-producing yeast Yarrowia lipolytica. Appl. Environ. Microbiol. 80, 1660–1669.

Mirończuk, A.M., Rzechonek, D.A., Biegalska, A., Rakicka, M., Dobrowolski, A., 2016. A novel strain of Yarrowia lipolytica as a platform for value-added product synthesis from glycerol. Biotechnology for Biofuels 9 (180). https://doi.org/10.1186/s13068-016-0593-z.

Mischo, H.E., Proudfoot, N.J., 2013. Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. Biochim. Biophys. Acta 1829, 174–185.

Mishra, P., Lee, N.-R., Lakshmanan, M., Kim, M., Kim, B.-G., Lee, D.-Y., 2018. Genome-scale model-driven strain design for dicarboxylic acid production in Yarrowia lipolytica. BMC Syst. Biol. 12, 12.

Moon, H.Y., Van, T.L., Cheon, S.A., Choo, J., Kim, J.-Y., Kang, H.A., 2013. Cell-surface expression of Aspergillus saitoi-derived functional α-1,2-mannosidase on Yarrowia lipolytica for glycan remodeling. J. Microbiol. 51, 506–514.

Morín, M., Monteoliva, L., Insenser, M., Gil, C., Domínguez, A., 2007. Proteomic analysis reveals metabolic changes during yeast to hypha transition in Yarrowia lipolytica. J. Mass Spectrom. 42, 1453–1462.

Morin, N., Cescut, J., Beopoulos, A., Lelandais, G., Le Berre, V., Uribelarrea, J.-L., Molina-Jouve, C., Nicaud, J.-M., 2011. Transcriptomic analyses during the transition from biomass production to lipid accumulation in the oleaginous yeast Yarrowia lipolytica. PLoS One 6, e27966.

Müller, S., Sandal, T., Kamp-Hansen, P., Dalbøge, H., 1998. Comparison of expression systems in the yeasts Saccharomyces cerevisiae, Hansenula polymorpha, Klyveromyces lactis, Schizosaccharomyces pombe and Yarrowia lipolytica. Cloning of two novel promoters from Yarrowia lipolytica. Yeast 14, 1267–1283.

Mumberg, D., Müller, R., Funk, M., 1995. Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. Gene 156, 119–122. https://doi.org/10.1016/0378-1119(95)00037-7.

Naumova, E., Naumov, G., Fournier, P., Nguyen, H.V., Gaillardin, C., 1993. Chromosomal polymorphism of the yeast Yarrowia lipolytica and related species: electrophoretic karyotyping and hybridization with cloned genes. Curr. Genet. 23, 450–454.

Neuvéglise, C., Marck, C., Gaillardin, C., 2011. The intronome of budding yeasts. Comptes

Rendus Biologies, Ten years of genomic exploration in eukaryotes : strategy and progress of Genolevures. 334. pp. 662–670. https://doi.org/10.1016/j.crvi.2011.05.015.

Nicaud, J.-M., 2012. Yarrowia lipolytica. Yeast 29, 409–418.

Nicaud, J.M., Fabre, E., Gaillardin, C., 1989a. Expression of invertase activity in Yarrowia lipolytica and its use as a selective marker. Curr. Genet. 16, 253–260.

Nicaud, J.M., Fabre, E., Gaillardin, C., 1989b. Expression of invertase activity in Yarrowia lipolytica and its use as a selective marker. Curr. Genet. 16, 253–260.

Nicaud, J.M., Fournier, P., La Bonnardière, C., Chasles, M., Gaillardin, C., 1991. Use of ars18 based vectors to increase protein production in Yarrowia lipolytica. J. Biotechnol. 19, 259–270.

Nicaud, J.M., Madzak, C., van den Broek, P., Gysler, C., Duboc, P., Niederberger, P., Gaillardin, C., 2002. Protein expression and secretion in the yeast Yarrowia lipolytica. FEMS Yeast Res. 2, 371–379.

Oberhardt, M.A., Palsson, B.Ø., Papin, J.A., 2009. Applications of genome-scale metabolic reconstructions. Mol. Syst. Biol. 5, 320.

Ogrydziak, D.M., Scharf, S.J., 1982. Alkaline extracellular protease produced by Saccharomycopsis lipolytica CX161-1B. J. Gen. Microbiol. 128, 1225–1234.

Otero, R.C., Cordero Otero, R., Gaillardin, C., 1996. Efficient selection of hygromycin-B-resistant Yarrowia lipolytica transformants. Appl. Microbiol. Biotechnol. 46, 143–148.

Pan, P., Hua, Q., 2012. Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, Yarrowia lipolytica. PLoS One 7, e51535.

Papanikolaou, S., Aggelis, G., 2003a. Modeling Lipid Accumulation and Degradation in Yarrowia lipolytica Cultivated on Industrial Fats. Curr. Microbiol. 46, 398–402.

Papanikolaou, S., Aggelis, G., 2003b. Modelling aspects of the biotechnological valorization of raw glycerol: production of citric acid byYarrowia lipolytica and 1,3-propanediol byClostridium butyricum. J. Chem. Technol. Biotechnol. 78, 542–547.

Papanikolaou, S., Galiotou-Panayotou, M., Chevalot, I., Komaitis, M., Marc, I., Aggelis, G., 2006. Influence of glucose and saturated free-fatty acid mixtures on citric acid and lipid production by Yarrowia lipolytica. Curr. Microbiol. 52, 134–142.

Park, Y.K., Dulermo, T., Ledesma-Amaro, R., Nicaud, J.M., 2018a. Optimization of odd chain fatty acid production by Yarrowia lipolytica. Biotechnology for Biofuels. 201811, 158. https://doi.org/10.1186/s13068-018-1154-4.

Park, Y.-K., Nicaud, J.-M., Ledesma-Amaro, R., 2018b. The Engineering potential of Rhodosporidium toruloides as a Workhorse for Biotechnological applications. Trends Biotechnol. 36, 304–317. https://doi.org/10.1016/j.tibtech.2017.10.013.

Pignede, G., Wang, H.-J., Fudalej, F., Seman, M., Gaillardin, C., Nicaud, J.-M., 2000. Autocloning and Amplification of LIP2 in Yarrowia lipolytica. Appl. Environ. Microbiol. 66, 3283–3289.

Pomraning, K.R., Kim, Y.-M., Nicora, C.D., Chu, R.K., Bredeweg, E.L., Purvine, S.O., Hu, D., Metz, T.O., Baker, S.E., 2016. Multi-omics analysis reveals regulators of the response to nitrogen limitation in Yarrowia lipolytica. BMC Genomics 17. https://doi.org/10.1186/s12864-016-2471-2.

Rakicka, M., Rywińska, A., Cybulski, K., Rymowicz, W., 2016. Enhanced production of erythritol and mannitol by Yarrowia lipolytica in media containing surfactants. Braz. J. Microbiol. 47, 417–423. https://doi.org/10.1016/j.bjm.2016.01.011.

Richard, G.-F., Kerrest, A., Lafontaine, I., Dujon, B., 2005. Comparative genomics of hemiascomycete yeasts: genes involved in DNA replication, repair, and recombination. Mol. Biol. Evol. 22, 1011–1023. https://doi.org/10.1093/molbev/msi083.

Rigouin, C., Gueroult, M., Croux, C., Dubois, G., Borsenberger, V., Barbe, S., Marty, A., Daboussi, F., André, I., Bordes, F., 2017. Production of Medium Chain Fatty Acids by Yarrowia lipolytica: Combining Molecular Design and TALEN to Engineer the Fatty Acid Synthase. ACS Synth. Biol. 6, 1870–1879. https://doi.org/10.1021/acssynbio.7b00034.

Robles-Rodriguez, C.E., Bideaux, C., Guillouet, S.E., Gorret, N., Cescut, J., Uribelarrea, J.-L., Molina-Jouve, C., Roux, G., Aceves-Lara, C.A., 2017. Dynamic metabolic modeling of lipid accumulation and citric acid production by Yarrowia lipolytica. Comput. Chem. Eng. 100, 139–152.

Rodriguez, G.M., Hussain, M.S., Gambill, L., Gao, D., Yaguchi, A., Blenner, M., 2016. Engineering xylose utilization in Yarrowia lipolytica by understanding its cryptic xylose pathway. Biotechnol. Biofuels 9, 149.

Sánchez, B.J., Zhang, C., Nilsson, A., Lahtvee, P.-J., Kerkhoven, E.J., Nielsen, J., 2017. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol. Syst. Biol. 13, 935. https://doi.org/10.15252/msb.20167411.

Sassi, H., Delvigne, F., Kar, T., Nicaud, J.-M., Coq, A.-M.C.-L., Steels, S., Fickers, P., 2016. Deciphering how LIP2 and POX2 promoters can optimally regulate recombinant protein production in the yeast Yarrowia lipolytica. Microb. Cell Factories 15, 159.

Schmid-Berger, N., Schmid, B., Barth, G., 1994. Ylt1, a highly repetitive retrotransposon in the genome of the dimorphic fungus Yarrowia lipolytica. J. Bacteriol. 176, 2477–2482.

Schwartz, C.M., Hussain, M.S., Blenner, M., Wheeldon, I., 2016. Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in Yarrowia lipolytica. ACS Synth. Biol. 5, 356–359.

Schwartz, C., Frogue, K., Ramesh, A., Misa, J., Wheeldon, I., 2017a. CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in Yarrowia lipolytica. Biotechnol. Bioeng. 114, 2896–2906.

Schwartz, C., Shabbir-Hussain, M., Frogue, K., Blenner, M., Wheeldon, I., 2017b. Standardized Markerless Gene Integration for Pathway Engineering in Yarrowia lipolytica. ACS Synth. Biol. 6, 402–409.

Schwartz, C., Curtis, N., Löbs, A.-K., Wheeldon, I., 2018. Multiplexed CRISPR Activation of Cryptic Sugar Metabolism Enables Yarrowia lipolytica Growth on Cellobiose. Biotechnol. J. 13 (9), e1700584. https://doi.org/10.1002/biot.201700584. Epub 2018 May 18.

Shabbir Hussain, M., Gambill, L., Smith, S., Blenner, M.A., 2016. Engineering Promoter

Architecture in Oleaginous yeast Yarrowia lipolytica. ACS Synth. Biol. 5, 213–223.

Shaw, A.J., Lam, F.H., Hamilton, M., Consiglio, A., MacEwen, K., Brevnova, E.E., Greenhagen, E., Latouf, W.G., South, C.R., van Dijken, H., Stephanopoulos, G., 2016. Metabolic engineering of microbial competitive advantage for industrial fermentation processes. Science 353, 583–586. https://doi.org/10.1126/science.aaf6159.

Sherman, D.J., Martin, T., Nikolski, M., Cayla, C., Souciet, J.-L., Durrens, P., Consortium, Génolevures, 2009. Génolevures: protein families and synteny among complete hemiascomycetous yeast proteomes and genomes. Nucleic Acids Res. 37, D550–D554.

Spagnuolo, M., Shabbir Hussain, M., Gambill, L., Blenner, M., 2018. Alternative Substrate Metabolism in Yarrowia lipolytica. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.01077.

Steensels, J., Snoek, T., Meersman, E., Nicolino, M.P., Voordeckers, K., Verstrepen, K.J., 2014. Improving industrial yeast strains: exploiting natural and artificial diversity. FEMS Microbiol. Rev. 38, 947–995. https://doi.org/10.1111/1574-6976.12073.

Swennen, D., Paul, M.-F., Vernis, L., Beckerich, J.-M., Fournier, A., Gaillardin, C., 2002. Secretion of active anti-Ras single-chain Fv antibody by the yeasts Yarrowia lipolytica and Kluyveromyces lactis. Microbiology (Reading, Engl.) 148, 41–50. https://doi.org/10.1099/00221287-148-1-41.

Swiat, M.A., Dashko, S., den Ridder, M., Wijsman, M., van der Oost, J., Daran, J.-M., Daran-Lapujade, P., 2017. FnCpf1: a novel and efficient genome editing tool for Saccharomyces cerevisiae. Nucleic Acids Res. 45, 12585–12598. https://doi.org/10.1093/nar/gkx1007.

Tai, M., Stephanopoulos, G., 2013. Engineering the push and pull of lipid biosynthesis in oleaginous yeast Yarrowia lipolytica for biofuel production. Metab. Eng. 15 (1–9). https://doi.org/10.1016/j.ymben.2012.08.007.

Thevenieau, F., Nicaud, J.-M., Gaillardin, C., 2009. Applications of the Non-Conventional Yeast Yarrowia Lipolytica. In: Satyanarayana, T., Kunze, G. (Eds.), Yeast Biotechnology: Diversity and Applications. Springer, Netherlands, pp. 589–613. https://doi.org/10.1007/978-1-4020-8292-4_26.

Tomaszewska, L., Rakicka, M., Rymowicz, W., Rywińska, A., 2014. A comparative study on glycerol metabolism to erythritol and citric acid in Yarrowia lipolytica yeast cells. FEMS Yeast Res. 14, 966–976. https://doi.org/10.1111/1567-1364.12184.

Trassaert, M., Vandermies, M., Carly, F., Denies, O., Thomas, S., Fickers, P., Nicaud, J.-M., 2017. New inducible promoter for gene expression and synthetic biology in Yarrowia lipolytica. Microb. Cell Factories 16, 141.

Trébulle, P., Nicaud, J.-M., Leplat, C., Elati, M., 2017. Inference and interrogation of a coregulatory network in the context of lipid accumulation in Yarrowia lipolytica. npj Systems Biology and Applications 3 (21). https://doi.org/10.1038/s41540-017-0024-1.

Tsakraklides, V., Kamineni, A., Consiglio, A.L., MacEwen, K., Friedlander, J., Blitzblau, H.G., Hamilton, M.A., Crabtree, D.V., Su, A., Afshar, J., Sullivan, J.E., Latouf, W.G., South, C.R., Greenhagen, E.H., Shaw, A.J., Brevnova, E.E., 2018. High-oleate yeast oil without polyunsaturated fatty acids. Biotechnol. Biofuels 11, 131.

Vandermies, M., Denies, O., Nicaud, J.-M., Fickers, P., 2017. EYK1 encoding erythrulose kinase as a catabolic selectable marker for genome editing in the non-conventional yeast Yarrowia lipolytica. J. Microbiol. Methods 139, 161–164.

Vandermies, M., Kar, T., Carly, F., Nicaud, J.-M., Delvigne, F., Fickers, P., 2018. Yarrowia lipolytica morphological mutant enables lasting in situ immobilization in bioreactor. Appl. Microbiol. Biotechnol. 102, 5473–5482. https://doi.org/10.1007/s00253-018-9006-5.

Verbeke, J., Beopoulos, A., Nicaud, J.-M., 2013. Efficient homologous recombination with short length flanking fragments in Ku70 deficient Yarrowia lipolytica strains. Biotechnol. Lett. 35, 571–576.

Wagner, J.M., Alper, H.S., 2016. Synthetic biology and molecular genetics in non-conventional yeasts: current tools and future advances. Fungal Genet. Biol. 89, 126–136.

Wagner, J.M., Williams, E.V., Alper, H.S., 2018. Developing a piggyBac Transposon System and Compatible selection Markers for Insertional Mutagenesis and Genome Engineering in Yarrowia lipolytica. Biotechnol. J. https://doi.org/10.1002/biot.201800022.

Wang, G., Li, D., Miao, Z., Zhang, S., Liang, W., Liu, L., 2018. Comparative transcriptome analysis reveals multiple functions for Mhy1p in lipid biosynthesis in the oleaginous yeast Yarrowia lipolytica. Biochim. Biophys. Acta 1863, 81–90. https://doi.org/10.1016/j.bbalip.2017.10.003.

Wasylenko, T.M., Ahn, W.S., Stephanopoulos, G., 2015. The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in Yarrowia lipolytica. Metab. Eng. 30, 27–39.

Wojtatowicz, M., Rymowicz, W., Academy of A, 1991. Effect of inoculum on kinetics and yield of citric acid production on glucose media by Yarrowia lipolytica A-101. Acta Aliment. Pol. Pol.

Wong, L., Engel, J., Jin, E., Holdridge, B., Xu, P., 2017. YaliBricks, a versatile genetic toolkit for streamlined and rapid pathway engineering in Yarrowia lipolytica. Metabolic Engineering Communications 5, 68–77.

Xie, D., 2017. Integrating Cellular and Bioprocess Engineering in the Non-conventional Yeast Yarrowia lipolytica for biodiesel production: a Review. Front Bioeng Biotechnol 5 (65). https://doi.org/10.3389/fbioe.2017.00065.

Xie, D., Jackson, E.N., Zhu, Q., 2015. Sustainable source of omega-3 eicosapentaenoic acid from metabolically engineered Yarrowia lipolytica: from fundamental research to commercial production. Appl. Microbiol. Biotechnol. 99, 1599–1610. https://doi.org/10.1007/s00253-014-6318-y.

Xue, Z., Zhu, Q.Q., 2012. Ammonium transporter promoter for gene ex-pression in oleaginous yeast. US Patent 8323960.

Xue, Z., Sharpe, P.L., Hong, S.-P., Yadav, N.S., Xie, D., Short, D.R., Damude, H.G., Rupert, R.A., Seip, J.E., Wang, J., Pollak, D.W., Bostick, M.W., Bosak, M.D., Macool, D.J., Hollerbach, D.H., Zhang, H., Arcilla, D.M., Bledsoe, S.A., Croker, K., McCord, E.F., Tyreus, B.D., Jackson, E.N., Zhu, Q., 2013. Production of omega-3 eicosapentaenoic

acid by metabolic engineering of Yarrowia lipolytica. Nat. Biotechnol. 31, 734–740.

Yang, X.-S., Jiang, Z.-B., Song, H.-T., Jiang, S.-J., Madzak, C., Ma, L.-X., 2009. Cell-surface display of the active mannanase in Yarrowia lipolytica with a novel surface-display system. Biotechnol. Appl. Biochem. 54, 171–176. https://doi.org/10.1042/BA20090222.

Yang, L.-B., Dai, X.-M., Zheng, Z.-Y., Zhu, L., Zhan, X.-B., Lin, C.-C., 2015. Proteomic Analysis of Erythritol-Producing Yarrowia lipolytica from Glycerol in Response to Osmotic pressure. J. Microbiol. Biotechnol. 25, 1056–1069.

Yovkova, V., Otto, C., Aurich, A., Mauersberger, S., Barth, G., 2014. Engineering the α-ketoglutarate overproduction from raw glycerol by overexpression of the genes encoding NADP + −dependent isocitrate dehydrogenase and pyruvate carboxylase in Yarrowia lipolytica. Appl. Microbiol. Biotechnol. 98, 2003–2013. https://doi.org/10.1007/s00253-013-5369-9.

Yue, L., Chi, Z., Wang, L., Liu, J., Madzak, C., Li, J., Wang, X., 2008. Construction of a new plasmid for surface display on cells of Yarrowia lipolytica. J. Microbiol. Methods 72, 116–123.

Yuzbasheva, E.Y., Yuzbashev, T.V., Laptev, I.A., Konstantinova, T.K., Sineoky, S.P., 2011. Efficient cell surface display of Lip2 lipase using C-domains of glycosylpho-sphatidylinositol-anchored cell wall proteins of Yarrowia lipolytica. Appl. Microbiol. Biotechnol. 91, 645–654.

Zhang, B., Rong, C., Chen, H., Song, Y., Zhang, H., Chen, W., 2012. De novo synthesis of trans-10, cis-12 conjugated linoleic acid in oleaginous yeast Yarrowia Lipolytica. Microb. Cell Factories 11, 51. https://doi.org/10.1186/1475-2859-11-51.

Zhang, B., Chen, H., Li, M., Gu, Z., Song, Y., Ratledge, C., Chen, Y.Q., Zhang, H., Chen, W., 2013. Genetic engineering of Yarrowia lipolytica for enhanced production of trans-10, cis-12 conjugated linoleic acid. Microb. Cell Factories 12, 70. https://doi.org/10.1186/1475-2859-12-70.

Zhang, J.-L., Peng, Y.-Z., Liu, D., Liu, H., Cao, Y.-X., Li, B.-Z., Li, C., Yuan, Y.-J., 2018. Gene repression via multiplex gRNA strategy in Y. lipolytica. Microb. Cell Factories 17, 62.

1.4   **Yeast factories for the production of aromatic compounds**

Aromatic chemicals have numerous applications in chemical, food, feed, polymer and pharmaceutical industries. A large proportion of relevant aromatic compounds are naturally produced by plants, which means that their production depends on plant life cycles, weather, plant disease, requires high amount of biomass, and faces difficult separation processes leading to low yields. As a result, since their biological production is still mostly inefficient or unavailable, most aromatic compounds and their derivatives are nowadays chemically synthesised from petroleum-derived sources via not sustainable procedures (Gottardi et al., 2017; Huccetogullari et al., 2019; Noda and Kondo, 2017).

Aromatic rings are not readily available in the environment, and the main biological route for their formation is the branched pathway to Tryptophan (Trp), Phenylalanine (Phe), and Tyrosine (Tyr) occurring in plants, fungi and bacteria (Lehninger et al., 2000). Thus, the aromatic amino acid (AAA) biosynthetic pathway is a major source of bio-based aromatic compounds (Brückner et al., 2018; Gottardi et al., 2017; Suástegui and Shao, 2016), and consequently an important target to engineer in order to transform microorganisms in cell factories for the production of commercially relevant aromatic compounds (Figure 4).

In fact, a large variety of aromatics compounds of interest such as hormones, nutraceuticals and dyes have already been synthesized by microbial cell factories. For instance, pinocembrin, used in food and pharmaceutical industries due to its anti-inflammatory, anti-microbial and anti-cancer activities was produced by *E. coli* and *S. cerevisiae* (Eichenberger et al., 2017; Leonard et al., 2008); vanillin, a widely used fragrance and aroma agent in various industries, was synthesized in *E. coli*, *Schizosaccharomyces pombe* and *S. cerevisiae* (Hansen et al., 2009; Overhage et al., 2003); *E. coli* and *Pseudomonas putida* were also engineered to produced anthranilic acid, which is used for the synthesis of dyestuff, perfumes and pharmaceuticals (Balderas-Hernández et al., 2009; Kuepper et al., 2015). The current status of these microbial production and the metabolic engineering strategies employed was recently reviewed by several groups (Gottardi et al., 2017; Huccetogullari et al., 2019; Suástegui and Shao, 2016).
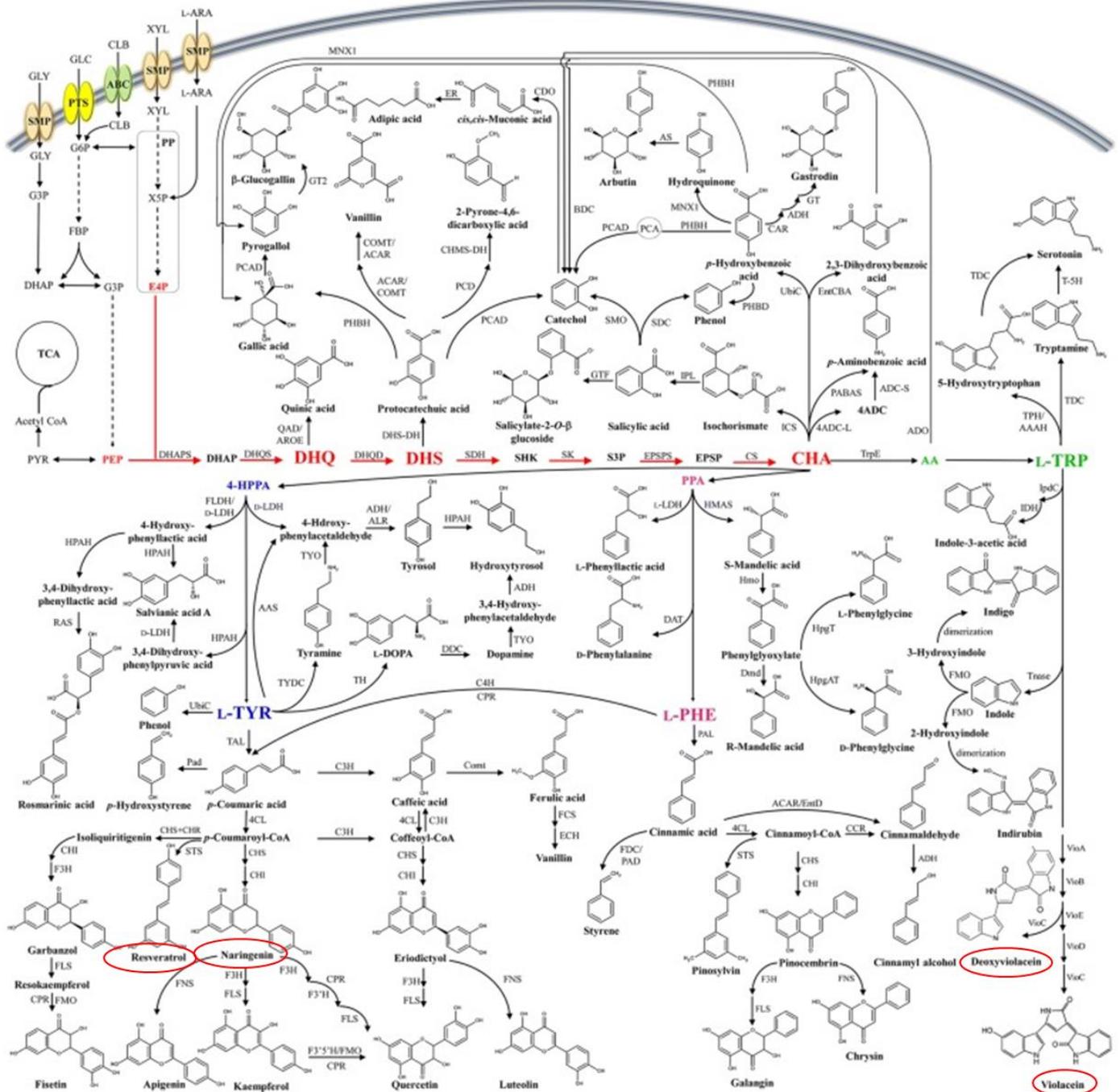
Adapted from Huccetogullari et al., 2019

Figure 4. Biosynthetic pathways of various aromatic compounds derived from the aromatic amino acid pathway. In red circles are highlighted molecules described in the text.

Despite the fact that *Y. lipolytica* has naturally high acetyl-CoA and malonyl-CoA pools (Christen and Sauer, 2011), which are necessary for the biosynthesis of the aromatic molecules such as flavonoids and stilbenes (see 1.4.5.3 and 1.4.5.4), the AAA pathway has not been very much studied nor engineered in this yeast. Celinska and co-

workers described the production of 2-phenylethanol (2PE), a product of Phe catabolism with rose-like aroma, in *Y. lipolytica* (Celińska et al., 2013) and also analyzed and identified proteins that were involved in its production (Celińska et al., 2019, 2015b). Very recently, and while this work was being developed, two research teams achieved the production of heterologous aromatic products in *Y. lipolytica*, such as naringenin and resveratrol (Lv et al., 2019b, 2019a; Palmer et al., 2020). They also introduce in a small extend modifications on the shikimate pathway with the purpose of increasing its flux and thus the production of the compounds of interest.

Along the items of this section, the AAA pathway in yeasts will be described as well as its regulation and some of the derivative product that can be obtained. As mentioned, not many studies about AAA pathway have been done on *Y. lipolytica* up to date. However, given the fact that this metabolic pathway is highly conserved across species (Braus, 1991) and that genes coding for proteins highly similar to those from well described organisms are found in *Y. lipolytica*, one can expect that the pathway is conserved and comparable to what it is described for other better-studied yeasts. Although differences in enzymes activities and their regulation might exist (Christen and Sauer, 2011), the descriptions of the pathway done here are mainly based on *S. cerevisiae* information.

1.4.1 The aromatic amino acid biosynthetic pathway

The AAA pathway can be sectioned into three main parts: (1) the shikimic acid pathway, (2) the L-Trp branch, and (3) the L-Tyr and L-Phe branches (Figure 5) (Braus, 1991; Suástegui and Shao, 2016).

The shikimate pathway, common to all AAA, is composed of seven enzyme-catalysed reactions encoded by four genes and leads to chorismate. The first committed step in the shikimate pathway is the condensation of erythrose 4-phophate (E4P) and phosphoenolpyruvate (PEP), catalysed by deoxy-d-arabino-heptulosonate-7-phosphate (DAHP) synthase, for which two isoenzymes exist, with different substrate affinity and feedback regulation (see 1.4.3), Aro3 and Aro4. The formed DAHP is then converted to chorismate by Aro1 and Aro2 sequentially. In yeasts, Aro1 is a multifunctional enzyme that catalyses 5 steps. Chorismate is the first branch point of the pathway, with one branch leading to Trp and the other to Phe and Tyr. The biosynthesis of Trp proceeds in five steps. The first step is the conversion of chorismate to anthranilate by the

anthranilate synthase (Trp2) and glutamine amidotransferase (Trp3). Anthranilate is successively transformed by enzymes Trp4, Trp1, Trp3 and Trp5 to Trp. On the other hand, only two further steps are needed, from chorismate, for the biosynthesis of Tyr and Phe. Chorismate mutase, encoded by *ARO7*, produces prephenate, which is then directed to either Tyr or Phe. Prephenate dehydrogenase, encoded by *TYR1*, transforms prephenate into *p*-hydroxyphenylpyruvate, while prephenate dehydratase, encoded by *PHA2*, transforms prephenate into 2-oxo acid phenylpyruvate. *p*-hydroxyphenylpy-ruvate and phenylpyruvate can then be transaminated to Tyr or Phe respectively by the aromatic amino acids transaminases I (Aro8) or II (Aro9) (Braus, 1991; Iraqui et al., 1998). A schematic representation of the AAA pathway is presented in Figure 5.
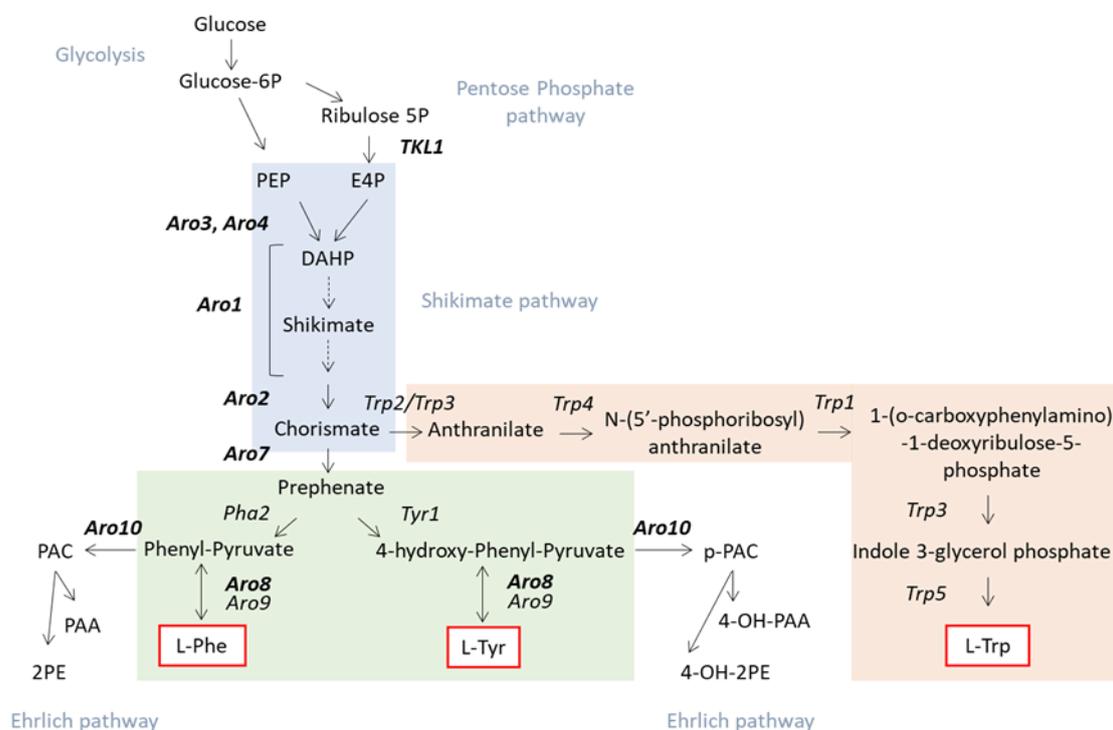
Figure 5. Aromatic amino acid pathway

Representation of the AAA pathway divided into three parts: (1) the shikimic acid pathway (blue), (2) the L-Trp branch (red), and (3) the L-Tyr and L-Phe branches (green). The precursor availability through glycolysis and pentose phosphate pathway as well as the catabolism through the Ehrlich pathway are shown. Genes involved in the pathway are indicated next to the arrow representing the reaction. Genes shown on bold letters were involved in the construction of engineered strains presented during this work. Phe: phenylalanine; Tyr: Tyrosine; Trp: Tryptophan. 2PE: phenylethanol; PAA: phenylacetic acid; 4-OH-2PE: 2-(4-hydroxyphenyl)ethanol; 4-OH-PAA: 4-hydroxyphenyla-cetic acid

1.4.2   **The Ehrlich pathway**

Several amino acids, the AAA, the branched-chain and the methionine, can be assimilated by the Ehrlich pathway and converted into higher fusel alcohols. The pathway consists of three reactions, (1) transamination of the amino acid, (2) decarboxylation of the 2-oxo-acid formed in the preceding reaction and (3) a reduction of the resulting fusel aldehyde into the higher fusel alcohol (Hazelwood et al., 2008) (Figure 5 and Figure 6).

In *S. cerevisiae* each step of the Ehrlich pathway can be conducted by several possible enzymes, however, in *Y. lipolytica* the pathway appears to be less redundant. Looking at the metabolism of Phe and Tyr, their deamination can be performed by Aro8/Aro9 or Bat1/Bat2 in *S. cerevisiae* while in *Y. lipolytica* only Aro8 seems to be responsible of this reaction. In the oleaginous yeast, Bat1/Bat2 seem to be only involved in the branched-chain amino acids metabolism (Celińska et al., 2019) and Aro9 was not detected in a proteome analysis done after feeding the cells with AAA, condition that produced 2PE (Celińska et al., 2015b). The phenylpyruvate and *p*-hydroxyphenylpyruvate, produced from the deamination step, are subsequently decarboxylated to generate a fusel aldehyde. The decarboxylation can be done, in *S. cerevisiae*, by Aro10 or the pyruvate decarboxylases (Pdc1/3/5), however, in *Y. lipolytica* a single homolog of Aro10 was identified. The last activity of the pathway is a reduction of the fusel aldehyde by an alcohol dehydrogenase. Dickinson and co-workers have shown that any of the alcohol dehydrogenases (Adh1–5 and Sfa1) encoded within *S. cerevisiae* genome can catalyze this final reaction (Dickinson et al., 2003). On the other hand, in *Y. lipolytica*, Celinska and co-workers observed high over-representation of a single dehydrogenase having broad-range substrate specificity (YALI0F24937p) when Phe was added to the media (Celińska et al., 2015b). However, studies have shown that the balance between the reduction and the oxidation of the fusel aldehyde strongly depends on the redox status of the cell (Vuralhan et al., 2005, 2003). Thus, the fusel aldehyde is either reduced to fusel alcohols, 2PE or 2-(4-hydroxyphenyl)ethanol (4OH2PE), or oxidized to fusel acids, 2-phenylacetate (PAA) or 2-(4-hydroxyphenyl)acetate (4OHPAA) (Figure 6).

In terms of physiology, the Ehrlich pathway is used to seize nitrogen from amino acids (Thorne, 1949). In terms of chemical industry, it leads to the formation of valuable chemical compounds of significant interest. Ehrlich metabolites constitute an important

fraction of aroma compounds used in various commodities. For instance, the 2PE produced from Phe is seek for its rose-like odour (see 1.4.5.1), while PAA produces a honey-like aroma, and the 3-methylbutanol derived from leucine has a fruity odour (Holt et al., 2019). In addition, 4OH2PE (tyrosol), has been described as heart protector due to its effect on inhibiting cholesterol oxidation in low density lipoproteins (LDL) (Caruso et al., 1999).
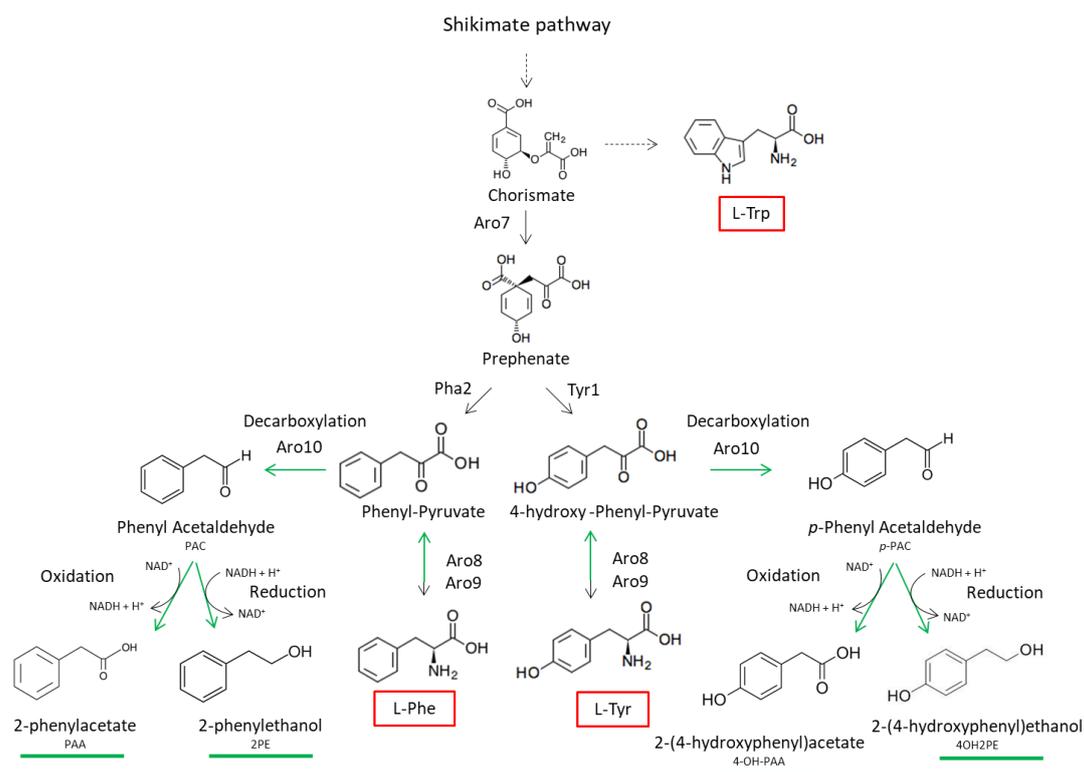
Figure 6. Ehrlich pathway

Representation of the Ehrlich pathway from Phe and Tyr (green arrows). In green lines are highlighted the four Ehrlich metabolites obtained. In red squares are shown the AAAs.

The Ehrlich pathway in *Y. lipolytica* has been characterised by Celinska and co-workers. They evidenced that the production of 2PE at detectable levels was strain dependent, as well as the amount produced (Celińska et al., 2013). They also did a proteomic analysis to identify enzymes involved in the pathway. By this analysis they found that Aro8 was constitutively expressed, while Aro10 was upregulated in samples were Phe was added into the media. On the other hand, the coding sequence of Aro80, a transcription factor required for the transcription of *ARO9* and *ARO10* in *S. cerevisiae*, was found in the genome of *Y. lipolytica* but it was not identified as expressed at the protein level (Celińska et al., 2015b). Finally, by engineering the Ehrlich pathway they

confirmed that Aro8 has an aminotransferase activity and plays a role in the deamination of Phe and that Aro10 is responsible for the decarboxylation activity (Celińska et al., 2019).

1.4.3   **Regulation of the aromatic amino-acid pathway**

Synthesizing AAA requires twice the ATP cost than for most of the other amino acids. Therefore, tight regulations at the allosteric and transcriptional levels are expected throughout the pathway (Braus, 1991; Suástegui and Shao, 2016).

In *S. cerevisiae*, the initial step of the shikimate pathway is regulated through feedback inhibition by end products of the pathway. Thus, Aro4 is allosterically regulated by Tyr and Aro3 by Phe. Similarly, at the first branch point, Aro7 is feedback inhibited by Tyr and activated by Trp (Braus, 1991). Luttik and co-workers did a quantitative analysis of the impact of feedback inhibition on AAA biosynthesis in *S. cerevisiae*, showing that the introduction of feedback insensitive alleles of Aro4 and Aro7 results in a 4.5-fold increase of the flux through the AAA biosynthetic pathway (Luttik et al., 2008). In another work, Brückner *et al*. showed that the overexpression of a feedback resistant variant of Aro3 leads to an enhanced flux into the shikimate pathway, evidenced by a 2.3 fold-increase of the detected product (PCA) (Brückner et al., 2018). Furthermore, with lower K_m values for E4P and PEP and higher K_i for its specific inhibitor, Aro3 can be regarded as the more effective enzyme (Brückner et al., 2018; Hartmann et al., 2003).

In addition, the activity through the pathway can be regulated by modifying the level of gene expression, thus, modifying the amount of enzyme produced. In this respect, certain transcription factors (TF) and metabolites were associated to transcription rates of genes in the AAA pathway (Braus, 1991). For instance, in *S. cerevisiae*, the TF Gcn4 is involved in transcription activation of amino acid biosynthesis genes, including *ARO1*, *ARO3*, *ARO4* and *ARO9*, in response to amino acids starvation (Braus, 1991; Ljungdahl and Daignan-Fornier, 2012). The TF Aro80, on its side, has a positive control effect on *ARO9* and *ARO10*, induced by Tyr, Phe and Trp (Iraqui et al., 1999).

Regarding the regulation of the pathway in *Y. lipolytica*, in a very recent Ph.D. thesis work carried out in our team by P. Trébulle, promoter regions of genes from the AAA pathway were analysed. This leads to the identification of several regulatory sequences conserved across Yarrowia clade: (i) a sequence associated to the TF Sfl1 was

identified on the promoter region of gene *TRP5*, (ii) The sequence associated to the binding of TF Gcn4, well-known in *S. cerevisiae* for being involved in amino acids metabolism, was also identified in genes of the shikimate pathway, and (iii) a sequence associated to the TF Dal80, which is part of the GATA TF family, was identified on the promoter region of *ARO3* and of several genes involved in the metabolism of amino acids such as *LEU3* (leucine synthesis), *ILV2* (for isoleucine and valine synthesis), *QNS1* (glutamine metabolism) and *ACO2* (involved in lysine synthesis), as well as in the promoter region of a gene coding for an amino acid permease (GAP1) (Trebulle, 2019). In addition, three other GATA TF are known to be conserved in *Y. lipolytica*: Gln3, Gat1 and Gzf3; these TF recognize a 5'-GATA-3' sequence and regulate genes subject to nitrogen catabolite repression (NCR) which is released under nitrogen limitation (Morin et al., 2011). Even though recent examination of genome-level transcriptional regulation on nitrogen limitation culture conditions has revealed that genes involved in amino acid biosynthesis are downregulated to increase the flux from carbon to lipid (Kerkhoven et al., 2016), further studies need to be conducted in order to better understand the regulation of amino acids in general, and AAA in particular, in *Y. lipolytica*. This knowledge will also be useful for further and more efficiently engineer the AAA pathway.

1.4.4   **Engineering the AAA biosynthetic pathway.**

With the intention to increase the flux through a pathway, four main strategies can be carried out: (i) overexpress genes involved in the pathway, (ii) change de regulation of the pathway, at the transcriptional level or at the allosteric level, (iii) eliminate or down regulate competing pathways and (iv) increase the availability of precursors.

S. cerevisiae has been extensively metabolically engineered to improve the production of AAA. Thus, the overexpression of *ARO1* and *ARO2* genes was applied to increase the levels of AAA derivate products such as 2PE and p-coumaric acid (Rodriguez et al., 2015). In addition to the overexpression of endogenous *ARO* genes, *AROL* from *E. coli.* was also expressed and showed positive effect on the flux through the pathway. AroL is one of the individual enzymes responsible of the transformation from DAHP to chorismate in this bacterium and was found to be the bottleneck in this six-reaction pathway (Hassing et al., 2019; Rodriguez et al., 2015). These examples evidence the effect of expressing genes involved in the engineered pathway.

The regulation of the AAA pathway was also modified in *S. cerevisiae.* The expression of the previously mentioned feedback insensitive forms of Aro4, Aro7 and Aro3 had a big impact on boosting the flux though the shikimate pathway. The co-expression of deregulated forms of Aro4^{K229L} and Aro7^{G141S} increased by 200-fold the extracellular concentration of aromatic fusel alcohols (Luttik et al., 2008). Thus, the strategy was also used to engineer a strain for the production of naringenin (Koopman et al., 2012). The combined expression of Aro4^{K229L} and Aro3^{K222L} increase on 3.3-fold the production of protocatechuic acid (Brückner et al., 2018). And recently, Aro4^{K229L} Aro3^{K222L} and Aro7^{G141S} were expressed together in an engineered strain to further increase the production of 2PE. (Hassing et al., 2019).

Regarding the elimination of competing pathways, knocking-out *ARO10* avoids the catabolism of Phe and Tyr by the Ehrlich pathway. Thus, the pool of these AAAs can be used, for instance, for the synthesis of *p*-coumaric acid or naringenin (Koopman et al., 2012; Rodriguez et al., 2015). In the case of competing pathways that are somehow essential for the cell survival, one could change the promoter of the gene in order to have a downregulated gene expression instead of deletion. This approach was, for instance, implemented in a *S. cerevisiae* strain constructed for the production of 2PE (Hassing et al., 2019). In that work, researchers decreased the flux through the Tyr branch by changing the promoter of *TYR1* by a very weak one, thus, prephenate flux was favoured to Phe, and subsequently to the production of 2PE, while avoiding the Tyr auxotrophic state seen when knocking out *TYR1*.

Finally, the precursors can be increased by modifying the culture composition favouring the presence of precursors in the medium, but also by engineering metabolic pathways that will lead to improved precursor supply (Lim et al., 2011; Pandey et al., 2016). In *S. cerevisiae*, the deletion of *ZWF1* together with the overexpression of *TKL1* increased the supply of E4P (Brückner et al., 2018; Hassing et al., 2019). It is to be noted that the deletion of *ZWF1* can have some effects in NADPH supply, thus limitations in product titers can be expected if NADPH demands of the pathways used are high. In this case, other sources could be overexpressed to compensate (Brückner et al., 2018). This is an example of how complex the metabolic connections and its regulations are in the cell, and evidence the challenge of engineering cell factories.

1.4.5   **Examples of compounds derived from the AAA pathway**

AAA, as source of natural aromatic structure, are the building block for the production of a large number of industrial interesting products by microorganism. A non-exhaustive list is presented in Figure 4. As example, five of these compounds are further described hereafter and were produced as proof of concept during this work.

1.4.5.1   **Violacein and protodeoxyviolaceinic acid**

Violacein is a purple pigment produced as secondary metabolite of microorganisms such as *Chromobacterium violaceum, Collimona*, *Duganella*, *Janthinobacterium* and *Pseudoalteromonas*. In addition to its use as a pigment in the cosmetic and fabric industry, this compound is known to have diverse biological activities and it is considered as potential drug agent in pharmaceutical industry due to its anti-microbial, anti-tumor and anti-oxidant activities. These characteristics have increase the interest on this compound, increasing efforts to better understand its mechanism of action and to improve its production yields (Choi et al., 2015; Durán et al., 2007).

In natural producers, violacein is synthesized from L-Trp via the *vioABCDE* operon that encodes VioA, VioB, VioE, VioD and VioC. Two molecules of L-Trp are condensed by VioA and VioB to form indole-3-pyruvic acid (IPA). Then, IPA is decarboxylated to form protodeoxyviolaceinic acid (PVA) via VioE. Subsequently, protodeoxyviolaceinate is oxidized to violacein by VioD and VioC. If VioC is expressed, without VioD, deoxyviolacein is generated, which lacks one hydroxyl group and also has interesting biological activities as violacein (Rodrigues et al., 2013; Sánchez et al., 2006) (Figure 7). In addition, the intermediate PVA has the characteristic of being a green compound and only three genes of the pathway are needed to synthesize it, which make it an interesting molecule to be used as a reporter molecule.
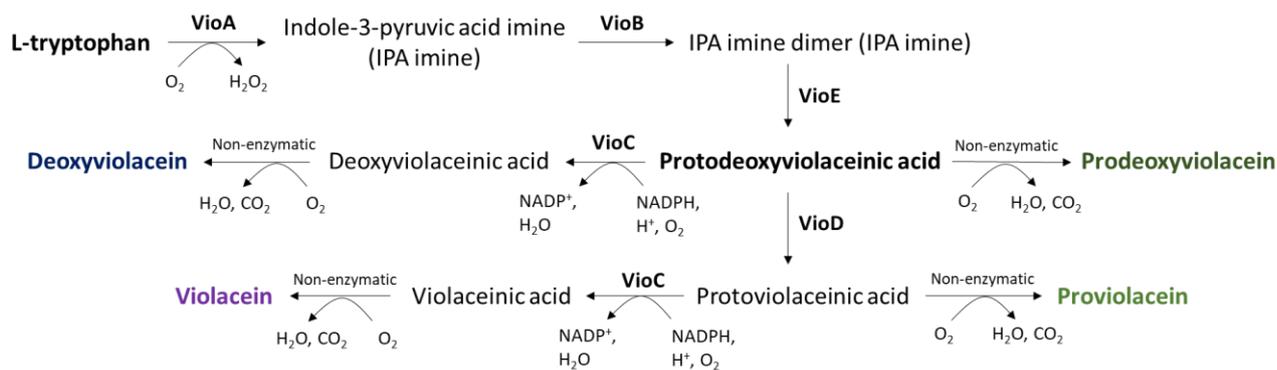
Figure 7. Violacein biosynthetic pathway

The enzymes catalysing the different reactions of the pathway, VioA, VioB, VioC, VioD and VioE, are indicated.

With the objective of increasing the production of this interesting compound, research have been conducted to improve the culture conditions, extraction and detection methods, and engineered strains have been constructed (Choi et al., 2015). Consequently, with metabolic engineering and fermentation optimization practice a production of 5.4 g/L of crude violacein (Violacein + deoxyviolacein) was achieved in *Corynebacterium glutamicum,* which is the highest production of this compound reported so far (Sun et al., 2016). Regarding *Y. lipolytica*, this pathway has been successfully expressed producing violacein in the order of mg/L (Kholany et al., 2019; Wong et al., 2017), however no codon optimization of the genes was done nor the strains were engineered to favour the availability of the precursor, Trp. Thus, efforts can be done to further increase the production of this compound in this oleaginous yeast.

1.4.5.2   2-Phenylethanol

2-Phenylethanol (2PE) is a valuable rose-like aroma compound widely used in cosmetics and food production, and can additionally be used as building block for other products (Etschmann et al., 2002). The global production was estimated to be 10,000 tons/year in 2010 (Hua and Xu, 2011) and its market was estimated at US$700 million in 2019 according to a report done by BCC Research (Pandal, 2014). Because of cost effectiveness, its actual production is based on chemical synthesis, with a price of about US$ 5/kg (Hua and Xu, 2011). However, due to the substantial negative environmental impact of these methods and the toxic by-products produced (Etschmann et al., 2002), efforts are being done to replace them by biotechnological production processes. Furthermore, the European and US regulations restricted food grade 2PE to natural sources which include botanical and microbiological sources, comprising fermentation products

(European parliament and Council 2008). The natural source from essential oils of flowers and plants has the inconvenient of producing low concentration of 2PE and, additionally, its purification is difficult (Mei et al., 2009), which is reflected in the final price of the natural extracted 2PE, about US\$ 1000/kg (Hua and Xu, 2011). In micro-organism, 2PE can be synthesised from Phe via the Ehrlich metabolism (Celińska et al., 2013; Etschmann et al., 2002) (Figure 8. See also 1.4.2). The advantage of biological synthesis is that, thanks to new techniques of synthetic biology, it can be easily engineered to increase the production.
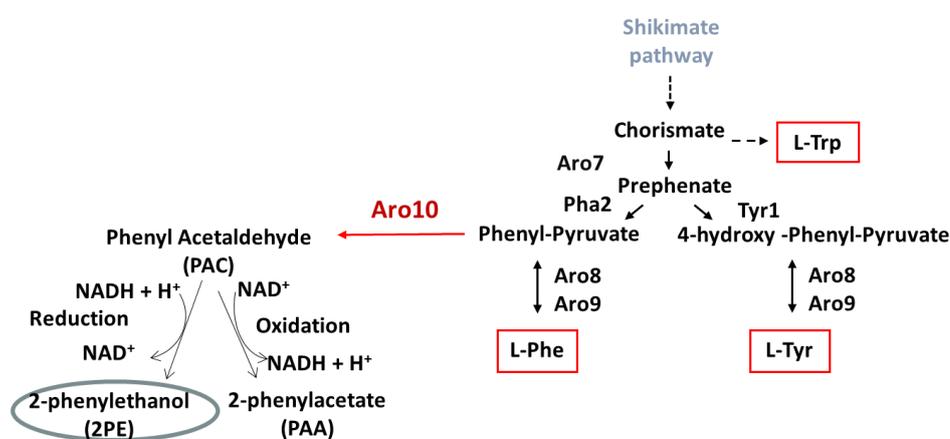
Figure 8. Biosynthesis pathway of 2-phenylethanol

Schematic representation of the biosynthetic pathways leading to 2PE from Ph

Celinska and co-workers provided the first evidence that *Y. lipolytica* could produce 2PE (Celińska et al., 2013). In their work, six different wild-type *Y. lipolytica* strains were cultured and analysed. In the supernatant of four of them 2PE was detected, indicating that the ability to produce of 2PE is strain-dependent. Using the best identified wild-type producer strain, a bioconversion test was done reaching 1.98 g/L 2PE by adding 7g/L Phe into the media at stationary growth phase. When using a strain over-expressing the Ehrlich pathway, the addition of 2g/L Phe in the medium lead to 0.65 g/L of 2PE (Celińska et al., 2019).

In Celinska's work the production of 2PE by *Y. lipolytica* was done by bioconversion from pure amino acids, which has an economic impact on the process. Metabolic engineering strategies may successfully redirect necessary carbon skeletons from central metabolism to be further converted by Ehrlich pathway activities. As an example,

an engineered *S. cerevisiae* strain reached a production of 1.6 g/L 2PE, which is the highest *de novo* production of this compound in yeasts to date (Hassing et al., 2019).

An important advantage of *Y. lipolytica* over *S. cerevisiae* is its Crabtree-negative status, which eliminates the risks of a synergic toxicity of ethanol and 2PE phenomenon that amplifies the toxicity of the later (H. Wang et al., 2011). Thus, *Y. lipolytica* appears as a better option for the production of this compound dedicated to human applications.

1.4.5.3   Naringenin

Naringenin is part of the flavonoid family, which are a large family of secondary plant metabolic intermediates with a C6–C3–C6 carbon framework as common core structure. Most of them are responsible for the colour of plant structures (Pandey et al., 2016). In humans, flavonoids are associated with a plethora of health benefits, the most remarkable ones being antioxidant, antimicrobial, anti-inflammatory, photo-protection, and reducer of heart disease and cancer (Nijveldt et al., 2001). The global market of flavonoids was estimated at USD 410.1 million in 2015 USD and is expected to increase up to 1.06 billion by 2025, according to a report by Grand View Research, Inc (Grand View Research, 2016b). Plant production of flavonoids is limited by the low production efficiency. To overcome this limitation, metabolic engineering and synthetic biology approaches have been applied to produce these compounds in heterologous microorganism such as *E. coli* and *S. cerevisiae* (Koopman et al., 2012; Y. Wang et al., 2011).

For its biosynthesis, Phe or Tyr, produced via shikimate, are the starting precursors. Tyrosine ammonia lyase (TAL) or phenylalanine ammonia lyase (PAL) converts tyrosine or phenylalanine to their carboxylic acid derivatives, *p*-coumaric acid and cinnamic acid, respectively. The conversion of cinnamic acid to *p*-coumaric acid requires an additional enzyme, cinnamic acid 4-hydroxylase (C4H). 4-coumaroyl-CoA ligase (4CL) converts p-coumaric acid into activated 4-coumaroyl-CoA, one of the precursors in C6–C3–C6 backbone biosynthesis of different flavonoid classes. Another precursor is malonyl-CoA, mainly produced via acetyl-CoA by the irreversible action of acetyl-CoA carboxylase (ACC) complex. Three molecules of malonyl-CoA and a molecule of 4-coumaroyl-CoA are condensed by the action of chalcone synthase (CHS) to form a C6–C3–C6 backbone unit (naringenin chalcone) in all flavonoids. Naringenin chalcone is converted into naringenin in a ring-closing step by chalcone isomerase (CHI).

Naringenin is itself a starting point for the biosynthesis of other flavonoids (Pandey et al., 2016) (Figure 9).
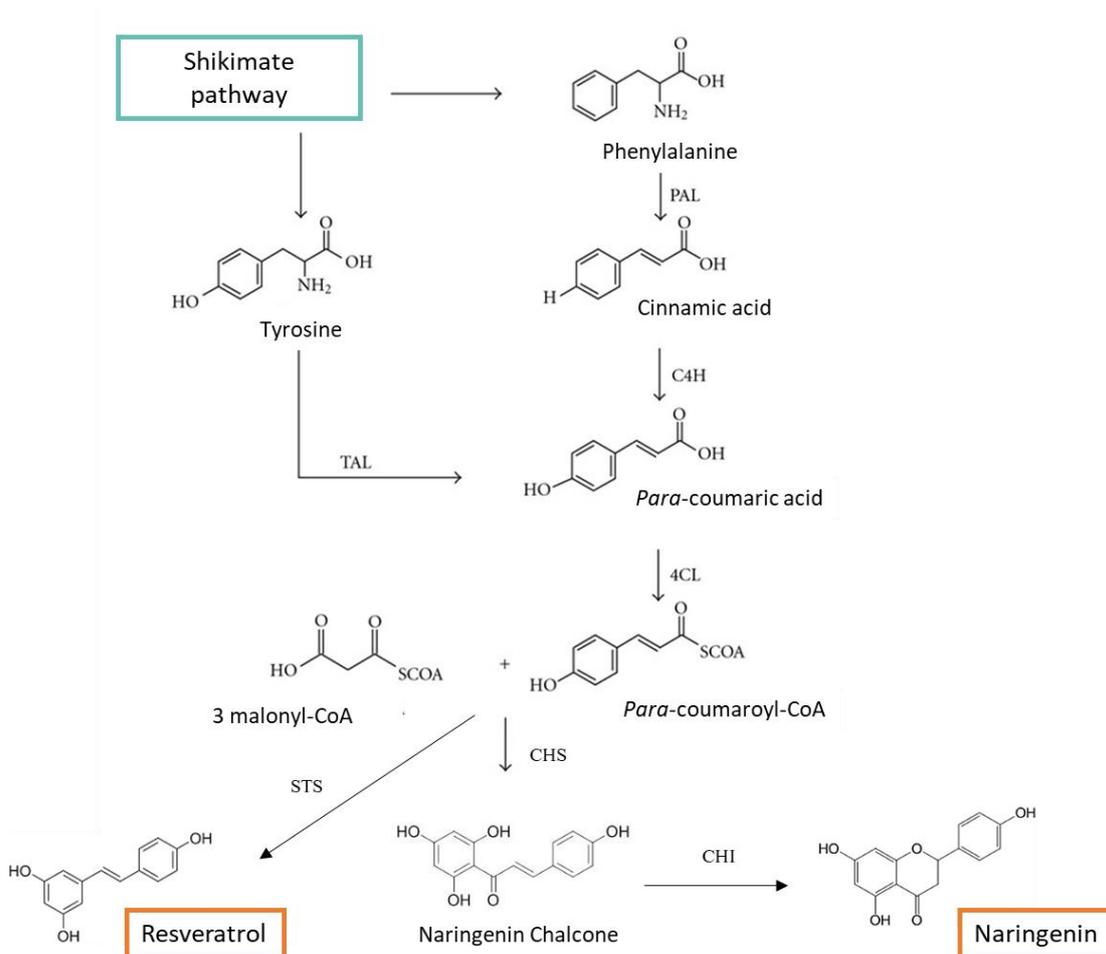
Figure 9. Naringenin and resveratrol biosynthetic pathways

Schematic representation of the biosynthetic pathways leading to naringenin and resveratrol from Phe and Tyr.

In *S. cerevisiae*, the highest amount of naringenin produced is of 100 mg/L in a strain engineered to have an improved AAA pathway (Koopman et al., 2012). Very recently, naringenin has been produced in *Y. lipolytica* (Lv et al., 2019b, 2019a; Palmer et al., 2020). Palmer *et al.* produced 124 mg/L in a *Y. lipolytica* engineered strain, and were able to scale up to 898 mg/L in 3 L bioreactor with optimized culture conditions, which is the highest reported titer of naringenin produced in a host cell (Palmer et al., 2020).

1.4.5.4 Resveratrol

Resveratrol, 3,5,4'-trihydroxystilbene, is a well-known polyphenol from the stilbene family produced as a secondary metabolite by plants, where it plays a role as defence against pathogens infection and injury (Bru et al., 2006). In humans many beneficial properties have been described for resveratrol, such as activities against various kinds of cardiovascular and nerve-related diseases, and also, anticancer, antiaging, antidiabetic activities (Pangeni et al., 2014). Its global market value was estimated at US$ 97.7 million in 2018 and is expected to grow by 8.1% by 2028, according to a report from Future Market Insight (Future Market Insights, 2018). However, as it is the case for many plants derived products, the low content, difficult extraction and purification methods and seasonal occurrence make the production of commercially amounts impracticable and unsustainable. On the other hand, the chemical synthesis of resveratrol produces high yields, but it is a technical complex synthesis and many unwanted byproducts are obtained. Therefore, the use of microbial host for heterologous resveratrol biosynthesis appears as a promising alternative (Mei et al., 2015) to face the increasing demand.

The biosynthetic pathway is highly similar to the one of naringenin. Precursors of resveratrol are either L-Phe or L-Tyr which are converted into *p*-coumaric acid, and afterwards into *p*-coumaroyl-CoA. Finally, catalyzed by stilbene synthase (STS), three molecules of malonyl-CoA are condensed with 4-coumaroyl-CoA to produce resveratrol (Thapa et al., 2019) (Figure 9).

Resveratrol has been synthesised in heterologous host such as *E. coli*, *Lactococcus lactis* and *S. cerevisiae*. The highest *de novo* production concentration achieved to date are 304 mg/L in *E. coli* and of 812 mg/L in fed batch culture of *S. cerevisiae* (Thapa et al., 2019). In *Y. lipolytica*, a first production was done and patented by DuPont (Huang et al., 2006), producing 1,46 mg/L. Very recently, Palmer and co-workers produced 8,8 mg/L of *de novo* resveratrol in a *Y. lipolytica* engineered strain (Palmer et al., 2020).

1.4.5.5 Melanin

Melanins comprise a heterogeneous group of polymeric pigments that are widely found in nature, from humans to microorganisms. They are the result of oxidation and polymerization of phenolic or indolic compounds. The resultant macromolecules have

high molecular weight, are negatively charged, hydrophobic and can vary in colour. They are known to be insoluble in organic solvents and acid aqueous solutions, and soluble in alkaline solutions (Plonka and Grabacka, 2006).

Based on their biosynthesis pathway, they were classified into three groups: eumelanins, pheomelanins and allomelanins. Eumelanins are the product of the oxidation of Tyr and/or Phe through *o*-dihydroxyphenylalanine (DOPA) intermediate, resulting in a polymer that displays a brown or black colour. Pheomelanins are initially synthesized like eumelanins but then DOPA undergoes cysteinylation, resulting in pigment with red-yellow colour. The allomelanins are the result of oxidation of either one of the following compounds: dihydroxynaphthalene (DHN), homogentisic acid (HGA), 4-hydroxyphenylacetic acid, catechols, γ-glutaminyl-4-hydroxybenzene or tetrahydroxynaphthalene, protocatechualdehyde, and caffeic acid. The oxidation of HGA, synthesised from Tyr, produces a brown pigment named pyomelanins (Plonka & Grabacka, 2006) (Figure 10).

Melanins can have diverse applications due to their chemical composition and physicochemical properties. These polymers can act as ultraviolet light, X-ray and γ-ray absorbers, reactive oxygen species (ROS) and free radical scavengers, and ion exchangers (Y. C. Liu et al., 2015). In addition to confer cells high protection against different environmental stresses, the benefits of the mentioned properties can be applied in a wide range of areas. Due to their UV protection and antioxidative effect they are incorporated in pharmaceutical and cosmetics preparations (ElObeid et al., 2017; Solano, 2017), and used for the production of contact lenses and sunglasses (Ahn et al., 2019). On the basis of their redox behaviour, melanins could be used as amorphous organic semiconductors in electronics (Bothma et al., 2008; Kim et al., 2013). Moreover, melanins could be used for bioremediation of contaminated sites due to their metal chelator capacity (Mahmoud, 2004). Melanin has been used to synthesize silver or gold nanostructures and nanoparticles, having potential uses in the food and health industries (Apte et al., 2013a; Kiran et al., 2014).
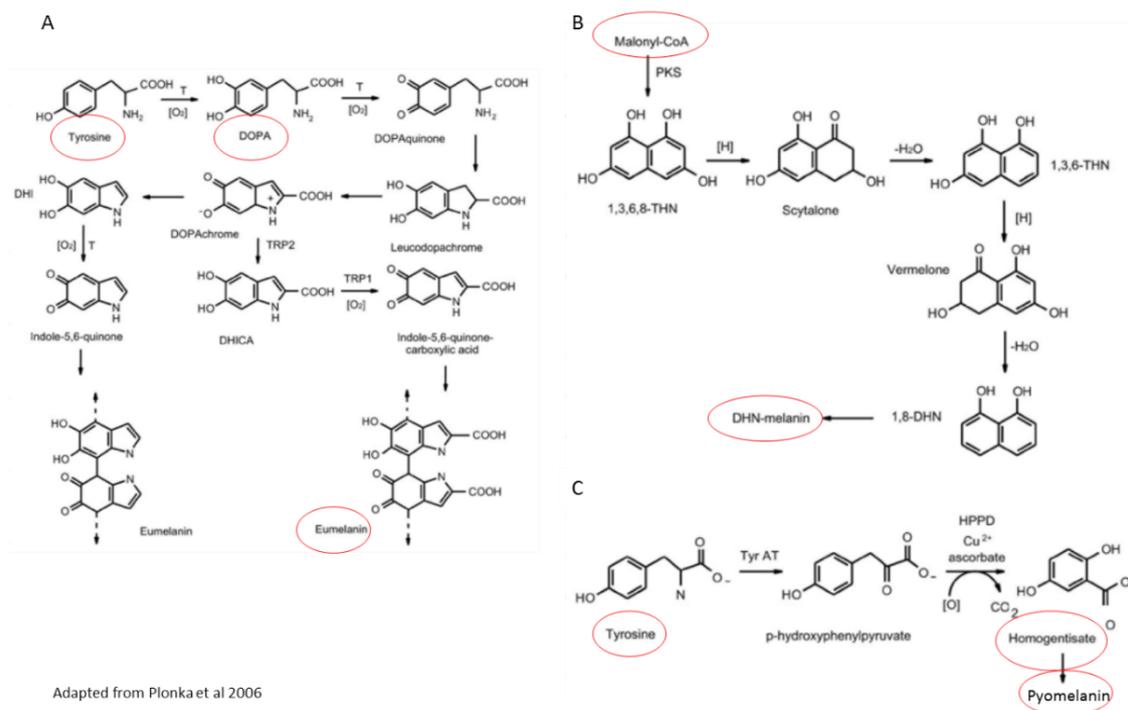
Figure 10. Melanin biosynthetic pathways.

A-Biosynthesis pathway of Eumelanin. B- Biosynthesis pathway of DHN melanin. C-Biosynthesis pathway of pyomelanin. In red circles are highlighted the precursors and products of each pathway.

Based on their useful properties, there is a growing global demand for melanins. Actually, melanins production by microorganisms is considered as an eco-friendly and economic approach (Martínez et al., 2019). Several recombinant microorganisms have been constructed for the production of melanin, being the higher production 28.8 g/L by *Streptomyces kathirae* (Guo et al., 2015).

The yeast *Y. lipolytica* is known for its ability to naturally produce melanin pigments from L-Tyr (Carreira et al., 2001a). This was first evidenced with the appearance of brown pigmentation during the ripening process of different types of cheese (Carreira et al., 1998), and later identified as a product of Tyr degradation. The pigment is produced in a biphasic process in which the pigment precursor, homogentisic acid (HGA), is first accumulated outside the cells and then autoxidized and polymerized, leading to the formation of pyomelanin (Carreira et al., 2001a; Carreira et al., 2001b). In addition, the culture conditions in which the pigment production was favoured, such as high aeration and neutral pH were identified (Carreira et al., 2001a). Some works have afterwards used *Y. lipolytica* to produce melanine with different purposes. Very recently, Ben Tahar *et al.* reported the production of pyomelanin by *Y. lipolytica* W29 wild-type

strain. In a Tyr-containing media, they achieved a yield of 0.5 g/L after 5 days of incubation and showed it antioxidant capacity and low cytotoxicity toward human keratinocytes, which make this compound interesting as UV filter in sunscreen (Ben Tahar et al., 2019). Ito *et al.* showed that melanin was constitutively synthetized in *Y. lipolytica* and that it contributes to sequester metal ions, giving *Y. lipolytica* a particular tolerance to cooper (Ito et al., 2007). In another series of works, Apte and co-workers enhanced the production of melanin by culturing cells of *Y. lipolytica* (NCIM 3590) in the presence of L-DOPA. Afterwards, due to the property of melanin of acting as an electron exchanger, they used the produced melanin for the synthesis of silver and gold nanoparticles. And finally described an application of the melanin-mediated silver nanoparticles as effective paint-additives displaying anti-fungal property (Apte et al., 2013a, 2013b).

Despite these specific features, little is still known on the nature and the mechanism of the pigment synthesis in *Y. lipolytica*. Thus, efforts can be addressed in order to increase the amount of melanin produced by *Y. lipolytica.*

As mentioned when introducing this sub-section (1.4.5), the hereinabove presented molecules are just five examples of the molecules that can be synthesised form the AAA pathway. Three works reviewing other molecules were cited when introducing in section 1.4. Among the interesting molecules that can be obtained from this pathway are, for instance: muconic acid and vanillin, both derived from DHS, an intermediate from the shikimate pathway. The first one serve as Nylon building block and the second one is used as flavouring compound in the industry. Neurotransmitters, serotonin and melanin can be synthesised from Trp, and L-DOPA from Tyr. Also, plant hormones as auxin are derived from this pathway.

1.5   Organization of the thesis

As previously stated, enlarging the scope of available cell factories is a key objective in order to increase bio-based sustainable processes. Therefore, the aim of this thesis is the construction of *Y. lipolytica* chassis strains, with optimized level of AAA that can be used as building blocks for the production of industrial relevant aromatic compounds. Furthermore, the development of new synthetic biology tools will ease the rapid and easy construction of cell factories.

Y. lipolytica, as described above, has proved to be a good candidate for cell factory. In addition, it is can be an interesting workhorse for the production of AAA-derived compounds due to its Crabtree-free effect, its strong endogenous flux through the TCA and PPP pathways and its high pool of malonyl-CoA.

After the description of the experimental methods used during this work in chapter **Material and methods** (2), the work will be divided into main parts. The chapter **Synthetic biology tools** (3) will describe the development of two modular synthetic biology tools, for DNA assembly and genome modification of *Y. lipolytica*. Their efficiencies and advantages will be discussed as well. Afterwards, the construction of engineered strains, using the developed tools, will be described in the chapter **Metabolic engineering of the aromatic amino acids pathway** (4). First, the sensor method used to detect changes on the engineered shikimate pathway will be presented. Then the construction of serial strains and the identification of the best AAA producer strain will be described. Finally, the use of the chassis strain for the biosynthesis of AAA-derived molecules of industrial interest will be shown. The regulation of the pathway, the production rates obtained with the constructed strains, the usefulness of the chassis strain and possible points to further increase the production with these strains are also analysed and discussed in this chapter.

2 Materials and methods

This chapter describes the general experimental methods and materials used to accomplish the objectives of the thesis.

2.1 Strains, media and growth conditions

Escherichia coli and *Y. lipolytica* strains used in this study are listed in section 6.1.3. They are stored at -80°C with 25 % glycerol for long-term conservation.

E. coli strain DH5α was used for cloning and plasmid propagation. Cells were grown at 37 °C with constant shaking on 5 mL Luria-Bertani (LB) medium (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl), and ampicillin (100 μg/mL) or kanamycin (50 μg/mL) were added for plasmid selection.

Y. lipolytica strains were grown at 28°C and constant shake (180 rpm) in different media depending on the pursued objective. Cultures in liquid media are preceded, except otherwise described, by a pre-culture of the fresh streaked cells strain for 16 h on 5 mL YPD which is afterwards centrifuged, washed with sterile water and inoculated at an initial OD of 0.05. When doing micro-cultures in 96-well plates, 200 μl of culture medium were inoculates with 2 μl of preculture. The preculture and the culture are done in in 200 μl of the same medium

The composition of media used for *Y. lipolytica* culture during this work are described hereafter.

Complete medium YPD (Yeast extract Peptone Dextrose): 1% yeast extract, 1 % BactoTM Peptone and 1 % glucose.

Minimal medium YNB (Yeast Nitrogen Based): 1.7 g/L YNB (without amino acids and ammonium sulfate), 5g/L NH$_4$Cl, 50mM phosphate buffer (pH 6.8), 10 g/L glucose.

Minimal medium YNB-Glc40: 1.7 g/L YNB, 5g/L NH$_4$Cl, 50mM phosphate buffer (pH 6.8), 40 g/L glucose.

Minimal medium YNB-Glc40-Decane10: 1.7 g/L YNB, 5g/L NH$_4$Cl, 50mM phosphate buffer (pH 6.8), 40 g/L glucose, 10% decane.

Minimal medium YNB-Gly40: 1.7 g/L YNB, 5g/L NH_4Cl, 50mM phosphate buffer (pH 6.8), 40 g/L glycerol.

When necessary, the YNB medium was supplemented with uracil (0.1 g/L), leucine (0.5 g/L) and/or lysine (0.8 g/L) and YPD with hygromycin (0.25 g/L) and/or Nourse-othricin (0.4 g/L). When needed they are referred in the text as YNBUra, YNBLeu, YNBHygro or YPDNour, respectively

In some cultures, when specified in during the text, Phe (1 g/L or 7g/L) or Tyr (1g/L) were added as supplement.

Solid media for *E. coli* and *Y. lipolytica* were prepared by adding 15 g/L agar (Invitrogen) to liquid media.

2.1.1 Specificities of culture conditions

2.1.1.1 Chassis strains

The series of AAA engineered strains constructed were cultured on 50 mL YNB in 500 mL baffled flasks, for five days. 1 mL was harvested after two- and five-days culture and used to measure growth by OD_{600} and evaluate compounds present in the supernatant by HPLC.

2.1.1.2 2PE producer strains

Strains were cultured on 50 mL of media, in 500 mL baffled flasks, for ten days. 1 mL was harvested after five- and ten-days culture and used to measure growth by OD_{600} and evaluate compounds present in the supernatant by HPLC.

When using decane in the medium, the sample was collected and measured after mixing it, in order to measure metabolites in both, aqueous and hydrophobic, phases.

2.1.1.3 Naringenin and resveratrol producer strains

Strains were cultured on 25 mL of media, in 250 mL baffled flasks, during five to twenty days. 1 mL was harvested every five days culture and used to measure growth by OD_{600} and compounds present in the supernatant by HPLC.

2.1.1.4   Melanin producer strains

The series of AAA engineered strains constructed were cultured on 15 mL YNB in 100 mL flasks for several days, up to 45 days. The progression of brown colour apparition was checked every two- three days by eye and documented.

2.2   Cloning and strain construction

2.2.1   General molecular biology

All restriction enzymes were purchased from New England Biolabs (NEB). PCR amplifications were performed using Q5 high-fidelity DNA polymerase (NEB) or GoTaq DNA polymerase (Promega). When needed, PCR fragment were purified using the QIAquick Gel Extraction Kit (Qiagen). Plasmids from *E. coli* were extracted using the QIAprep Spin Miniprep Kit (Qiagen). All the reactions were performed according to the manufacturer instructions.

2.2.2   Gene synthesis

Endogenous genes to be overexpressed were amplified from *Y. lipolytica* W29 genome by PCR and adapted according to our GG assembly protocol. Site-directed mutagenesis were done by a two-step PCR (see below) for elimination of internal BsaI recognition sites and for deregulation of Aro4 and Aro7 (section 3.2.2). Primers used for genome amplifications and site directed mutagenesis are listed in section 0.

Heterologous genes were codon optimized according to *Y. lipolytica* codon usage bias and synthesizes chemically by TWIST Bioscience (https://www.twistbioscience.com/). Codon optimizations were done either using the online software COOL (http://cool.syncti.org/) or the software provided by TWIST Bioscience interface. In order to make them compatible with our GG assembly strategy internal BsaI and BsmBI sites were avoided, and the corresponding external sites for GG assembly were added (see section 2.7). Ideally, the recognition sites of BamHI, ClaI, AvrII and IsceI enzymes are also removed from the optimized sequences, as they can be used for cloning in JMP vectors (see later). The sequences of these genes after codon optimization are shown in section 6.1.2.

2.2.2.1   Site-directed mutation by PCR

The DNA sequence to be modified is amplified in n+1 fragments, being n the number of mutations to do. To do so, divergent amplifications are done from the site to be mutated using the original sequence as template. Mutation is introduced in the divergent internal primers (Figure 11). Afterwards, the fragments obtained are put together and the final DNA is amplified using external primers (Primer1-Fw and Primer2-Rv in Figure 8) (Figure 11). The final PCR product should be treated with DpnI before use, to eliminate native template.
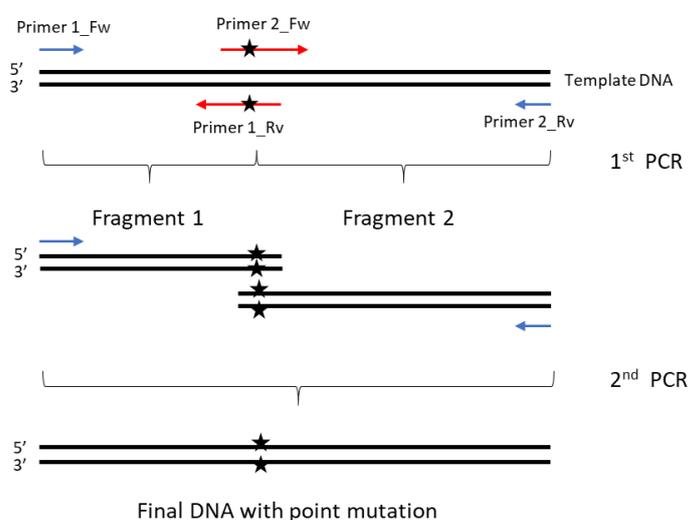
Figure 11. Point mutation PCR

Representation of a two-step PCR for introducing a point mutation into DNA.
The star represents the point mutation to be introduce in the DNA.

The PCR are performed using Q5 high-fidelity DNA polymerase (NEB). When assembling fragments (2nd PCR), 0,5 µl of each of the first PCR are added as template.

The design of the primers is critical. The divergent primers should have ~20 bp overlap, and each primer should have 10-15 nt at each side of the mutation point. The primers are in general around 30-35 nt.

2.2.3   Expression cassette and vector construction

Expression cassettes for insertion into *Y. lipolytica* genome mostly contains, a transcription unit (promoter-gene-terminator), a gene for selection of transformants and sequences for insertion into the genome. A description of these parts was discuss in section 1.310

During this work, the assembly of the expression cassettes were mostly carried out by GoldenGate (GG) assembly technique (See 2.7). This GG method was also used to construct replicative plasmids for Cas enzymes expression.

Eventually, classical digestion-ligation method was used for cassette construction. Traditionally, expression vectors for *Y. lipolytica* were constructed by classical digestion and ligation method. JPM62 and JPM63 pre-assembled plasmids (Nicaud et al., 2002) containing zeta sequences, URA3 or LEU2 markers, pTEF and Tlip, are mostly used. The gene of interest can be inserted after digestion by BamHI and AvrII. In addition, marker can be changed by IsceI, and promoter by ClaI-BamHI digestion enzymes.

The expression cassettes are released by NotI digestion prior to *Y. lipolytica* transformation.

All the vectors constructed during this work are listed in the Annexe section 0

2.2.4 **Strategies for genome editing.**

Insertion of expression cassettes into the genome were done using the zeta elements, for random integration, or using 1kb homologous recombination sequences when a specific locus was targeted. The cassettes are released by NotI digestion from the vector prior to *Y. lipolytica* transformation

During this work, specific targeted genes disruptions were mostly done using CRISPR-Cas9 system (See 2.8).

The classical method for gene disruption in *Y. lipolytica* is based on a disruption cassette composed of an expression cassette flanked by 1Kb of homologous recombination sequences for the targeted gene (Fickers et al., 2003).

2.2.5 **Transformation**

2.2.5.1 *E. coli*. **heat shock transformation**

Chemically competent DH5α cells are taken out the -80 °C freezer and thaw on ice for 30 min. 5-10 µl of DNA plasmid are added to 80 µl of competent cell on a 1.5 mL tube, and the mix is kept on ice for 30 min. After a heat shock at 42°C for 30 s, 250 mL of LB medium are added to the mix and cultured for 1 h at 37 °C with constant shaking.

Cells are then plated on selective media and stored at 37 °C for 1 day until the appearance of colonies.

2.2.5.2 *Y. lipolytica.* **Lithium acetate (LiAc) transformation method**

Transformation of *Y. lipolytica* was performed using the lithium-acetate method adapted from (Barth and Gaillardin, 1996). *Y. lipolytica* strains are streaked on YPD plates and grown for 16 h at 28 °C. For each transformation a loopful of cells are scraped from the agar surface and wash in 1 mL sterile TE Buffer. After 1 min centrifugation at 10000 rpm, cells are resuspended in 600 µl LiAc 0.1M pH 6.0 and incubated 1 h in a water-bath at 28 °C. Cells are then centrifuged for 2 min at 3000 rpm and resuspended in 60 µl of LiAc 0.1 M pH 6.0. Afterwards, 40 µl of competent cells are transferred into a 2 mL tube, mixed with 3 µl of carrier DNA (5 mg/mL, Dualsystems AG Biotech. See S.3) and 10 µl of linearized vector (~ 200 ng), and incubated 15 min at 28 °C in a water-bath. Following, 350 µl of PEG solution (PolyEthylene Glycol in 0.1 M LiAc pH6) are added the transformation mix. After 1 h of incubation at 28 °C, a 10 min heat shock in a water-bath at 39 °C is performed. 600 µl of LiAc solution is added and cells are then plated on selective media. Transformants are selected on YNB, YNBLeu, YNBUra, YNBHygro or YPDNour media, depending on their genotype.

2.3   Verification of constructions

2.3.1   *E. coli* colony PCR

For verification of plasmids constructed, colony PCR was performed on 5 to 8 colonies according to the following protocol:

A PCR master mix for x reactions is prepared on ice (total volume of 25 µL per reaction):

Component	Volume (µL)
Water	(x+2) times 11.5 µL
1st forward primer	(x+2) times 0.5 µL
2nd reverse primer	(x+2) times 0.5 µL
2xGoTaq mix (Promega)	(x+2) times 12.5 µL

25 uL of the PCR master mix are dispensed into individual PCR tubes. A very small amount of single colonies are picked with a tip, transferred to the master mix PCR tube, and the following PCR protocol is run:

5 min at 95 °C; 30 x (30 s at 95 °C; 30 s at X °C; Y min at 72 °C); 5 min at 72 °C; Hold at 10 °C.

X correspond to the annealing temperature, which needs to be calculated for each pair of primers (https://france.promega.com/en/resources/tools/biomath/). Y is the time needed to amplify the desired sequenced, allowing 1 min/Kb.

10 µL of each reaction used for verification by electrophoresis in an agarose gel at 0.8% with ethidium bromide. DNA is visualized afterwards with UV-light and the obtained profile analysed.

2.3.2 Digestion by restriction enzymes

Plasmids, extracted from *E. coli* after overnight culture on selective media, are cut by restriction enzymes. Time, temperature of incubation and buffer for the reaction to happen are specific for each enzyme and should be verify on the specification sheet. After the incubation, the reaction is verified by electrophoresis in an agarose gel at 0.8% with ethidium bromide and the fragments size profile obtained is analysed.

2.3.3 *Y. lipolytica* colony PCR

For verification of strains constructed, colony PCR was performed on 4 to 8 colonies of *Y. lipolytica* with the following protocol:

Single, isolated colonies are picked with a 2 µL pipette tip or a toothpick and briefly dipped into a PCR tube containing 2 µL of sterile water. The following lysis protocol is run on a thermocycler:

30 seconds at 65°C; 30 seconds at 8°C; 1 minute 30 seconds at 65°C; 3 minutes at 97°C; 1 minute at 8°C; 3 minutes at 65°C; 1 minute at 97°C; 1 minute at 65°C; hold at 80°C (15 minutes)

In the meantime, the following PCR master mix for x reactions is prepared on ice (total volume of 20 µL per reaction):

Component	Volume (µL)
Water	(x+2) times 6 µL
1st forward primer	(x+2) times 1 µL
2nd reverse primer	(x+2) times 1 µL

2xGoTaq mix (Promega)   (x+2) times 10 µL

18 µL of the PCR master mix are added to each tube containing lysed yeast colonies and the following PCR protocol is run:

6 minutes initial denaturation at 94ºC, (35 s 94 °C, 45 s 60 °C, 2.5 min 68 °C) x 8 cycles with annealing temperature decreasing by 1ºC at every cycle; (35 s 94 °C, 45 s 52 °C, 2.5 min 68 °C)x32; 10 minutes final extension at 68ºC; Hold at 10 ºC

5-10 µL of each reaction for verification by electrophoresis (agarose gel 0.8% + ethidium bromide)

2.4 Analytical techniques

2.4.1 HPLC

An Agilent Zorbax Eclipse plus C18 Column (4.6 × 100, 3.5 micron) operating at 40°C was used. A gradient of acetonitrile and 20 mM KH2PO4 (pH 2) with 1% acetonitrile was used as eluent, at a flow rate of 0.8 mL·min^{-1}, increasing from 0 to 10% acetonitrile in 6 min followed by an increase to 40% acetonitrile until 23 min. From 23 min to 27 min, the flow was set to 99% KH2PO4. Standards were obtained from Sigma-Aldrich.

Culture supernatants were treated with 1 % (vol/vol) trifluoroacetic acid for at least one hour at 4°C, then centrifuged and filtrated (0.22 µm), before HPLC analysis.

Using a diode array detector (DAD), Ehrlich metabolites (PAA, 2PE, 4OH2PE and 4OHPAA), AAA (Phe, Tyr and Trp) and naringenin were measured at 200 nm and coumaric acid and resveratrol at 310 nm. Calibration curves were established for each of the compounds using commercial standards (purchase from Sigma) for each of them. Single compound solutions, in their specific solvent according to manufacturer instructions, of 10 mM were done for each compound and then diluted to final concentrations of 0.1, 0.2, 0.5, 1, 2 mM for HPLC calibration.

2.4.2 Absorbance and fluorescence measurements.

Optical density (OD$_{600}$) was measured at 600 nm using a Spectrophotometer to evaluate cell growth of flask cultures.

2.4.2.1 Growth and florescence analysis in 96 well plates

Y. lipolytica precultures were grown for 24 h in YNB medium (supplemented with Uracil or Leucine when needed) in 96-well plates. 2 µL were then transferred into 200 µL of fresh medium in 96-well microplates, obtaining a starting OD_{600nm} of around 0.1. YNB medium, supplemented if needed, was used for the growth and florescence analysis depending on the construction. Growth was performed and monitored in a microtiter plate reader Synergy Mx (Biotek, Colmar, France) following the manufacturer's instructions at 28 °C and constant agitation. OD_{600} and red, yellow and blue fluorescence were measured, depending on the construction, every 30 minutes for 72 hrs. Red fluorescence was analyzed with the wavelength settings ex: 558 nm/em: 586 nm, yellow: ex: 505nm/em: 530 nm and blue: ex: 435 nm/em: 478 nm. Fluorescence was expressed as mean specific fluorescence value (SFU/h, mean value of SFU per hours). Cultures were performed at least in duplicates.

Microplate AAA biosensor test

After preculture in YNB, 2 µL were transferred into a new plate with 198 µL YNB supplemented with 1, 0.5 or 0.1 mg/mL of Phe, Tyr or Trp. The plate was cultured at 28°C with constant shake on the plate and growth (OD_{600}) and fluorescence (Excitation: 558, Emission: 586) were measured every 30 min during 48 h. The test was done on duplicate on at least 3 clones of each construction.

2.4.2.2 PVA extraction and absorbance.

8 clones of each construction were cultured in a 96-well plate in 200 µL YNB. The OD_{600} and the PVA absorbance at 640 nm were measured during 72 h.

For the extraction of PVA, 150 µL of culture were added to 50 µL ethanol 98% in a new plate. The mix was incubated with constant shake for 24h. Centrifugation the supernatants were transferred into a new plate and the absorbance measure at 640nm.

2.4.2.3 Melanin extraction

5 mL of cultures were harvested, centrifuged and pellet was eliminated. Supernatant was acidified with HCl to pH~2. After 4 hours at room temperature samples were

centrifuged and a brown pellet was obtained, which was dried and weighed using a thermobalance WPS 110S (Radwag).

2.4.2.4 Naringenin biosensor

For the analysis of *Y. lipolytica* strains, 100 µl of an overnight preculture of *E. coli* containing the biosensor system (Trabelsi et al., 2018) was added into a 96-well plate together with 100 µl of the *Y. lipolytica* culture that is being done for naringenin production. The mixture was incubated for 5 h at 37°C with constant shake in a Biotek plate reader, and red fluorescence of the biosensor system was measured every 15 minutes.

A calibration curve was established using a naringenin standard from Sigma. YPD and YNB media were tested and none of them interfere with the red fluorescence measured.

2.5   qRT-PCR

Cultures were grown in 10 mL YPD in 100 mL baffled Erlenmeyer flasks at 28°C and 160 rpm. The inoculation was done at 0.2 OD using an overnight preculture. Cells were harvested 6 h post inoculation, frozen in liquid nitrogen, and stored at -80°C. RNA was extracted from the cells using the RNeasy Mini Kit (Qiagen, Hilden, Germany). Samples were quantified using Nanodrop, and 5 µg were treated with DNAse (TURBO DNA-Free kit, Invitrogen). cDNA synthesis and qPCR were done in one step using Luna Universal One-Step RT-qPCR Kit (NEB). Expression levels were performed through ΔCT and $\Delta \Delta CT$ methods.

$\Delta CT = CT_{(target\ gene)} - CT_{(calibrator\ gene)}$
$\Delta \Delta CT = [CT_{(target\ gene)} - CT_{(calibrator\ gene)}]_{target\ sample} - [CT_{(target\ gene)} - CT_{(calibrator\ gene)}]_{reference\ sample}$

Expression level $= 2^{-\Delta CT}$
Fold Change $= 2^{-\Delta \Delta CT}$

Primers were designed using the Primer3 program, respecting the following parameters:

- ✓ Tm ~60C (+/-2C)
- ✓ Amplicon ~150 pb (+/-20pb)
- ✓ Primer length ~20 bp (+/- 2 bp)

- ✓ Start with G or C at 5' for each primer, if possible
- ✓ Avoid 3' self-complementation for each primer (allow max 3bp)
- ✓ Avoid, if possible, Max self-complementation between primer pair
- ✓ Amplicons must be designed in 3' region of the gene, especially with sequences bigger than 1 Kb

For primers validation, the qPCR is done using genomic DNA instead of RNA in the same conditions than the samples.

- Primers are diluted 1/10 from 100mM stock.
- High purity gDNA should be used. Quantify it with Nanodrop.
- Dilutions of gDNA are made from 100ng, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}
- For each primer pair and each dilution, qPCR reaction are done at least in duplicate. It is important to consider a blank with H_2O instead DNA for each primer pair to evaluate background of amplification due to dimer formation or DNA contamination.
- Actin was used as endogenous reference gene (calibrator gene).
- CT average values against dilutions are graphed and a linear regression is performed. The slope from the regression formula is used to calculate efficiency.

  Efficiency = [[10 $^{(-1/m)}$]-1]X100 m=slope= -3.3 is considered as an efficiency of 100% Desired amplification efficiencies range from 90% to 110%. Otherwise, new primer pair need to be designed and tested again.

2.6   Enzymatic assays

The activities of chorismate mutase (EC 5.4.99.5) and DAHP synthase (EC 2.5.1.54) were tested according to (Luttik et al., 2008).

2.6.1   Protein extraction

Cells were cultured on 50 mL culture (YPD) overnight at 28 °C and constant shaking (150 rpm). Culture was completely collected, centrifuged for 5 minutes at 4000 rpm, and the pellet was washed with 25 mL TE buffer 1x. Cells were centrifuged again, resuspended on 1 mL of cold PBS 1x with antiproteases (cOmplete antiprotease inhibitor cocktail from Roche), and then divided into 5 vials with screw tops (0.5 mL in each tube). 0.5 mL of glass bead were added into each tube and after four runs of 20 s in a

Fast-Prep at speed setting 6 with intermittent cooling, samples were centrifuged for 3 min at 4000 rpm and finally the supernatants were collected into a new tube.

For DAHP-synthase assay, cell-extract should be dialysed to remove lower-molecular weight compounds. The Amicon Ultra 0.5 mL UltraCel-10k from Milipore (Ref: UFC501096) was used according to the manufacturer instructions.

2.6.2   Protein determination

The total amount of proteins was determined using the Pierce BCA Protein Assay kit from ThermoScientific (N° 23225), a colorimetric assay based on the reduction of Cu^{2+} to Cu^{1+} by proteins in an alkaline medium (biuret reaction). The absorbance of the colored solution is measured at 562 nm. Albumin is used as standard.

2.6.3   Chorismate mutase activity

To measure chorismate mutase activity, 1 mL incubation mix with 100 µL Tris–HCl pH 7.6 (50 mM), 10 µL dithiothreitol (DTT, 1 mM), and 10 µL EDTA (0.1 mM) was incubated with 100 µL cell extract at 30 °C. In order to test inhibition or activation of ARO7, either 0.5 mM of Tyr or 0.5 mM of Trp were used. At t = 0 min, 1mM (final concentration) of barium chorismate was added. During 5 min, a 100 µL sample was taken every minute, and added to 100 µL of 1 M HCl. These samples were incubated at 30 °C for 10 min, after which 800 µL of 1 M NaOH was added. Absorbance was measured at 320 nm, and an extinction coefficient of 13.165mM^{-1} cm^{-1} was used to calculate the phenylpyruvate concentration.

2.6.4   DAHP synthase activity

The DAHP-synthase assay was performed as follow: the mixture of 0.9 mL contained 100 µL phosphate buffer (100 mM), 100 µL PEP (0.5 mM), 100 µL E4P (1 mM) and H_2O to complete. The reaction was started with the addition of 100 µL cell extract. After 2, 3, 4 and 5 min of reaction time, 200 µL of the reaction mixture was transferred to an Eppendorf tube containing 80 µL of 10 % w/v trichloroacetic acid solution. The resulting mixture was centrifuged for 5 min in an Eppendorf table centrifuge to remove proteins. About 125 µL of the supernatants was transferred to clean Eppendorf tubes and 125 µL periodic acid (25 mM in 0.075 M H_2SO_4) was added to each tube. In order to check allosteric control of ARO4, either 0.5 mM of Tyr, 0.5 mM Phe or 0.5 mM of Trp were used. After incubation at room temperature for 45 min 250 µL of 2 % (w/v) sodium arsenite in 0.5 M HCl was added to destroy excess periodate (2 min at room

temperature). One millilitre of thiobarbituric acid (0.3 %, w/v, pH 2.0) was added to the tubes, which were subsequently placed in a boiling-water bath for 5 min. The mixture was cooled in a water bath for 10 min at 40 °C and the pink colour developed was measured at 549 nm in a spectrophotometer against air. The extinction coefficient of DAHP at 549 nm was 13.6 mM^{-1} cm^{-1}

2.7 **Golden Gate**

The GG strategy, used to assemble parts of DNA, is divided in two main steps. The first one consists in designing and synthesizing the parts that will be then assembled, while the second part is the assembly itself. As the setup of this technique was part of the tools development of his work, the principle of the technique as well as the construction of a library of parts that can be reused on different constructions and its characterisation, are further discuss in section 3.1.1.

2.7.1 **Construction of donor vectors with GG parts**

a. The DNA to be used for the GG parts should be screened for the presence of internal BsaI sites. If there are any they should be eliminated.

b. BsaI recognition sites together with site-specific 4nt should be added at each end of the DNA part. The specific 4nt are listed in Table 1, together with a description of primers to be used if the sites are added by PCR.

c. Designed DNA parts are cloned into the donor vector with a backbone bearing the kanamycin-resistance gene (e.g., Zero Blunt® TOPO® PCR Cloning Kit (ThermoFisher)) and transformed into *E. coli* competent cells.

d. Clones are verified by colony PCR and sequencing before storage at -80°C and further use.

Table 1. Description of the 4nt overhangs designed for the Y. lipolytica GG strategy and of the primers design to amplify the parts and add the BsaI site and the corresponding overhang

Position	Fw primer 5'-3'	Rv primer 5'-3'	5' overhang	3' overhang
InsertUp (NotI)	ggg<u>GGTCTC</u>t**GCCT**GCGG*CCGC*nnn	ccc<u>GGTCTC</u>t**ACCT**nnn	**GCCT**	**AGGT**
Marker	ggg<u>GGTCTC</u>t**AGGT**nnn	ccc<u>GGTCTC</u>t**CCGT**nnn	**AGGT**	**ACGG**
Promoter1	ggg<u>GGTCTC</u>t**ACGG**nnn	ccc<u>GGTCTC</u>t**CATT**nnn	**ACGG**	**AATG**
Gene1	ggg<u>GGTCTC</u>t**AATG**nnn	ccc<u>GGTCTC</u>t**TAGA**nnn	**AATG**	**TCTA**
Terminator1	ggg<u>GGTCTC</u>t**TCTA**nnn	ccc<u>GGTCTC</u>t**AAGC**nnn	**TCTA**	**GCTT**
Promoter2	ggg<u>GGTCTC</u>t**GCTT**nnn	ccc<u>GGTCTC</u>t**TTGT**nnn	**GCTT**	**ACAA**
Gene2	ggg<u>GGTCTC</u>t**ACAA**nnn	ccc<u>GGTCTC</u>t**ATCC**nnn	**ACAA**	**GGAT**
Terminator2	ggg<u>GGTCTC</u>t**GGAT**nnn	ccc<u>GGTCTC</u>t**TGAC**nnn	**GGAT**	**GTCA**
Promoter3	ggg<u>GGTCTC</u>t**GTCA**nnn	ccc<u>GGTCTC</u>t**GTGG**nnn	**GTCA**	**CCAC**
Gene3	ggg<u>GGTCTC</u>t**CCAC**nnn	ccc<u>GGTCTC</u>t**ATAC**nnn	**CCAC**	**GTAT**
Terminator3	ggg<u>GGTCTC</u>t**GTAT**nnn	ccc<u>GGTCTC</u>t**ACTC**nnn	**GTAT**	**GAGT**
InsertDown (NotI)	ggg<u>GGTCTC</u>t**GAGT**nnn	ccc<u>GGTCTC</u>t**CG-CAGC***GGCCGC*nnn	**GAGT**	**TGCG**
Destination vector	ggg<u>GGTCTC</u>t**TGCG**nnn	ccc<u>GGTCTC</u>t**AGGC**nnn	**TGCG**	**GCCT**

<u>Underlined sequences</u>: BsaI recognition site. **Bold sequences**: overhangs generated after digestion. *Italic sequences*: NotI recognition site. n: corresponds to the overlap of at least 20 nt with the sequence to be amplified. ggg and ccc are added before the recognition site of the enzyme in order to increase its efficacy.

2.7.1   **Assembly of expression vectors**

a.   GG parts (donor vectors) and the destination vector (GGE029 or GGE114 with ampicillin-resistance gene) are extracted from the *E. coli* using a commercial Miniprep Kit. The concentration of all plasmid preparations are determined using a Nanodrop.

b.   The GG assembly reaction is prepared and placed in the thermocycler:

All the GG parts needed and the destination vector are mixed in equimolar quantities (50 pmol). Add 1 µl BsaI + 1 µl T4 ligase + 2 µl T4 ligase buffer + up to 20 µl of ddH$_2$O.

Thermal program:

(37°C for 5 min; 16°C for 5 min) x 50; 37°C for 10 min; 80°C for 5 min; 15°C ∞

c.   To improve efficiency when assembling **three TUs**, the procedure can be split into two parts: the creation of preassembly constructs and the assembly of the multigene construct.

<u>Preassembly</u>:

A- Prepare three separate reactions with four GG parts each. All parts are represented in equimolar quantities (50 pmoles).

1- InsertionSiteUp + Marker + Promoter1 + Gene1

2- Terminator1 + Promoter2 + Gene2 + Terminator2

3- Promoter3 + Gene3 + Terminator3 + InsertionSiteDown.

B- Add to each reaction 0.5 µl BsaI + 0.5 µl T4 ligase + 1 µl T4 ligase buffer + up to 10 µl of ddH$_2$O.

C- Place each reaction in a thermocycler, and run the following thermal program:

(37°C for 3 min, 16°C for 2 min) x 30; 55°C for 5 min; 80°C for 5 min; 15°C ∞

Assembly of multigene construct:

A- Mix together the three previous reactions (1,2,3) in the same tube, and add the following:

10 µl Rxn1 + 10 µl Rxn2 + 10 µl Rxn3 + 50 pmoles destination vector + 2 µl BsaI + 2 µl T4 ligase + 4 µl T4 ligase buffer + up to 40 µl of ddH$_2$O

B- Place the mixture in a thermocycler, and run the following thermal program:

(37°C for 5 min, 16°C for 5 min) x 50; 55°C for 5 min; 80°C for 5 min; 15°C∞

d. Transform 10 µl of the assembly reaction (15 µl when the 2-step procedure is performed) to 80 µl of competent cells, and then plate the cells on LB agar + 100 µg/mL ampicillin.

e. White colonies (clones containing vectors from which the chromophore reporter protein (RFP) was released during the assembly process. See section 3.1.1) are verified by colony PCR. And after extraction, plasmids are also verified by NotI digestion.

2.7.2 **Transformation into a *Y. lipolytica* strain**

The plasmid of interest is extracted using a Miniprep commercial kit, then digested with NotI, and finally transformed into *Y. lipolytica* using the lithium-acetate method.

2.8 **CRISPR**

As well as GG, CRISPR-Cas9 was set up as part of this work making it a routine method for gene disruption in *Y. lipolytica* in our team. For this reason, further discussion on this technique are presented in section 3.1.2. Here I describe the protocol for cloning a double strand gRNA into the constructed CRISPR-Cas9 vectors using GG technique with the BsmBI enzyme.

2.8.1 Design gRNA molecules and relevant primers

a. Design guideRNA (gRNA) molecules with the tool CRISPOR (http://crispor.tefor.net/). Choose the sequence that target the gene of interest, preferably in the beginning-middle section of the open reading frame, with high efficiency score and the smaller possible number of predicted off-target.

b. The target sequences should be 20 nt in length and must not include the PAM sequence.

c. Design of primers harbouring gRNA for cloning into the plasmid:
GGP_gRNA_BsmbI_xxx_Fw

5'TTCGATTCCGGGTCGGCGCAGGTTGxxxxxxxxxxxxxxxxxxxxxGTTTTA 3'

GGP_gRNA_BsmbI_xxx_Rv:
5'GCTCTAAAACxxxxxxxxxxxxxxxxxxxxxCAACCTGCGCCGACCCGGAAT 3'

x represents the 20 nt target sequence. Letters highlighted in grey show the 4 nt overhang generated after BsmBI digestion.

2.8.2 Construction of the double stranded gRNA insert

a. Dissolve the complementary gRNA insert primers in distilled water to a final concentration of 100 μM.

b. Phosphorylate and anneal the primers with the following reaction:

Component	Volume (μL)
T4 Kinase	1 μL
Forward primer (100 μM)	1 μL
Reverse primer (100 μM)	1 μL
T4 Ligase buffer 10 X	1 μL
H2O	6 μL

c. Incubate at 37 °C for 30 minutes and then heat the mixture to 95 °C for 5 min. Finally ramp down to 25 °C at a 5 °C min^{-1} rate

d. Dilute primer mixture 1:200 with water

2.8.3 Cloning double stranded gRNA inserts into the CRISPR-Cas9 acceptor plasmid pGGA_CRISPRyl

a. Assemble the plasmid by GG assembly as follows:

Component	Volume (μL)
gRNA insert	2 μl
pGGA_CRISPRyl	100 ng
T4 Ligase buffer	2 μl
BsmBI	1 μl
T7 Ligase	1 μl
H$_2$O	to 20 μl

Incubate in a thermocycler with the following program:

(55°C for 5 min, 16°C for 5 min) x 30; 50°C for 5 min; 80°C for 5 min; 15°C ∞

b. Transform 10 μL of the reaction product into *E. coli*. Plate onto LB – ampicillin medium.

2.8.4 Check for correct assembly of the gRNA plasmids

a. Check 5 white colonies (clones containing vectors from which the chromophore reporter protein (RFP) was released during the assembly process. See section 3.1.1 and 3.1.2) by colony PCR (2.3.1) using the following primers:
    Fw Primer: VerifsgRNA_Fw5'CTTTGAAAAATACCTCTAATGCGCC 3'
    Rv Primer: VerifsgRNA_Rv5'AAGCACCGACTCGGTGCCA3'

b. Load 5 μL of the PCR product into an electrophoresis agarose gel 1%. If the incorporation of the gRNA molecule was successful, a 194 bp fragment should be visualised.

c. The assembly can also be verified by sequencing using the same primers.

2.8.5 Transform CRISPR-Cas plasmids to edit *Y. lipolytica* genome

The verified plasmid with the desired gRNA is transformed into *Y. lipolytica* using the lithium-acetate method. 200 μl of the transformation product are cultured in 10 mL of selective media for 24-48 h. The culture is then diluted in order to have 50-100 colonies per plate, plated on YPD and incubate at 28 °C. Colonies should appear after 2 days.

2.8.6 Check DNA mutation.

Check 5 colonies. DNA is first amplified by colony PCR (SEE 2.3.3)

a. Purify the PCR product using a commercial kit (e.g. NucleoSpin® Gel and PCR Clean-Up. Macherey-Nagel) and send samples for sequencing according to manufacturer instructions. Also, genomic DNA can be isolated and use directly for sequencing.   (*)

b. Align sequencing reads against reference sequence to identify mutations in the target sequence.

c. For long-term storage at -80°C, verified mutants are grown over night on YPD medium and then glycerol is added at 25%.

(*) before sending for sequencing a T7 endonuclease assay can be done in order to screen for mutations into the genome.

2.8.6.1 T7 endonuclease assay

This enzyme recognizes and cleaves non-perfectly matched DNA. Thus, it can be used to evidence mutations by annealing mutated and wild-type (or parental) DNA

strands. After amplification and annealing of DNAs, the enzyme is added. It will recognize the mismatched region and cut the DNA. The obtained fragments of DNA are evaluated by electrophoresis

- Total length of target PCR product should be 500 bp to about 1 kb,
- Primer should anneal, so that the distance of each primer to the expected cutting site is unequal (e.g. 300 bp and 700 bp), but at least 200 bp (makes running the gel and distinguishing T7 cleavage products easier).

a. Amplification by PCR

Do a first colony PCR to amplify the genomic DNA to test. Afterwards, use 0.35 µl of the colony PCR product as template for this second PCR.
This second PCR is necessary in order to increase the amount of DNA obtained.

50 µl Q5 (or Gotaq) PCR Mix: 0.35 µL DNA; 2,5 µL Fw Primer 1:10; 2,5 µL Rev Primer 1:10; 25 µL 2x Q5 reaction buffer; 19 µL H2O

98 °C for 30 s; (98 °C for 10 s, 64 °C for 30 s, 64 °C for 30 s) x 35; 72 °C for 2 min; 4 °C ∞

b. T7 assay
- Purify PCR product through gel extraction. Elute DNA with 30 µl H2O.
- Transfer 7,5 µl of each PCR DNA (7,5 µg "mutated DNA" + 7,5 µl "WT DNA". ~200-400 ng total DNA) into new tubes and add 1µl NEB buffer 2 (10x)
- Denature for 5 min at 95°C, then: switch off heating block and let samples cool down slowly for at least 25 min in the heating block
    →This step is for generating the heteroduplex DNA recognized by the T7 endonuclease.
- Add 0,3 µl T7 endonuclease (~3U)
- Incubate at 37°C for 1h30 min; then add Gel loading dye with EDTA (EDTA inactivates the T7 endonuclease)
- Run a 2% agarose gel at 100V for 40-60 min.

2.9   Lugol test

Lugol stain protocol was used to evidence *GSY* deletion.

Lugol's reagent (2% KI, 1% I2 in water; dissolved 1:1) is used for glycogen staining. When preparing Lugol's, reagent dissolve I2 in 5-10% of the final volume using KI solution. Only when it is completely dissolved add the rest of the water.

Lugol staining can be performed either on agar plate, either on microtiter plates
On agar plate:

- Transformants are replica plated and tested after 2 days with Lugol's (the colonies should not be too old because stationary phase cells deplete their glycogen stores).

- 3-4 mL of the solution are pipetted onto the plate. Pipette slowly to avoid colonies to disconnect from the plate and float around.

- The difference between negatives and positives is visible within 1-2 minutes (glycogen-positive (dark brown) and glycogen-negative (white or yellow) colonies). See Figure 12-A

- Some of the glycogen-negative colonies are picked from the original plate and the frame shift in *GSY* is confirmed by sequencing

<u>On 96-well cell culture plate</u>

- Isolated transformants are grown on 200 µl YPD for 12 h (one colony per well). A very small amount of colony is taken with a pipette-tip.

- Centrifuge the plate 5 min at 570 g. Eliminate the supernatant by quickly inverting the plate. (Make sure you have a duplicate of the plate. The plate should be duplicated unless you numbered the colonies in the original plate, and you know exactly which colony is in each well).

- 30 µl of lugol solution are pipetted onto the plate.

- The difference between negatives and positives is visible within 1-2 minutes, See Figure 12-B
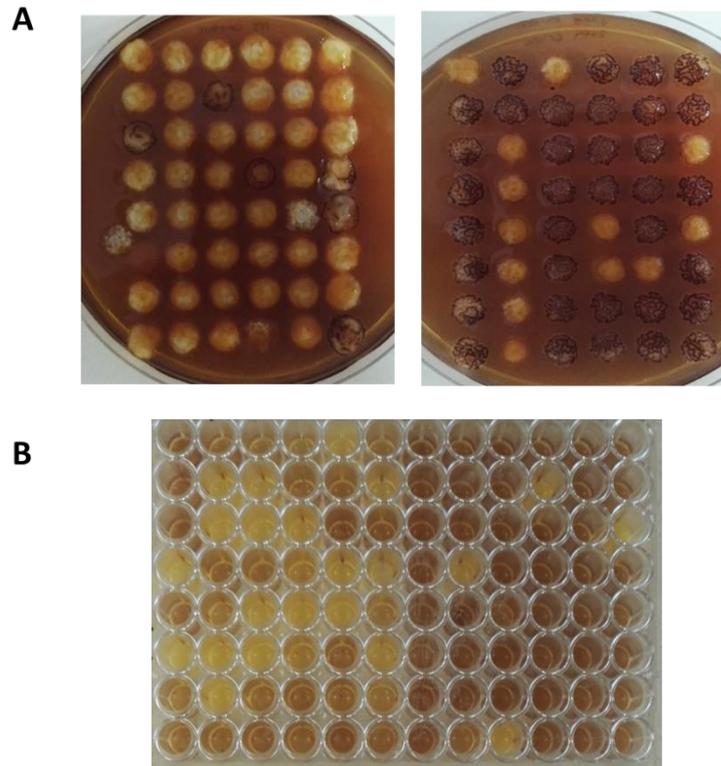
Figure 12. Screening of *GSY* gene disruption.

In A, the results of a petri plate test can be seen, while B shows the results of a well plate test. Dark brown are strains where the gene *GSY* is not disrupted and thus glycogen is synthesised and stain by Lugol solution. On the other hand, yellowish strains have the *GSY* gene disrupted so no glycogen can be produced and Lugol does not dye the cells.

3 Results

3.1 Synthetic biology tools

As already introduced in section 1.3, synthetic biology tools are methods, techniques and technologies that aim to engineer biological systems in a predictable and designed fashion (Young and Alper, 2010). The advent of new synthetic biology tools, highly efficient and robust, has accelerated and expanded the uses of metabolic engineering and improved the construction of cell factories.

Efforts done during the last years resulted in synthetic biology tools that are efficient, robust, versatile, fast, easy to use and easy to transfer to different organisms. Among the large number of tools that have been developed, several were applied in *Y. lipolytica*. This expansion of the toolbox is of great importance to further increase the range of applications for this biotechnological important yeast, which in turn raise the community interest on this yeast and favour the development of new tools and industrial strains. A work reviewing the most important synthetic biology tools developed for *Y. lipolytica* to date was done and presented in section 1.3.

During this thesis project, with the aim of expanding the toolbox dedicated for *Y. lipolytica,* a DNA assembly technique, named Golden Gate (GG), and a genome editing technique using CRISPR-Cas9, were set-up for their used in this yeast. The results obtained are presented hereunder.

3.1.1 Golden Gate assembly

The DNA assembly is the centrepiece of synthetic biology, allowing the construction of novel biological systems and devices by physically linking DNA fragments. Thus, the development of methods enabling DNA assembly in a faster and more efficient manner is a priority task for the easier construction of strains with complex genetic functionalities.

GG is a modular cloning approach, designed to assemble multiple genes via a single-step, one-pot reaction (Engler et al., 2008) (Figure 13). The system is based on type IIS restriction enzymes, which cut outside their recognition sites and produce an overhang that can be, thus, arbitrarily defined. The system is designed in such a way that

the overhangs generated allow the assembly of DNA fragments in a defined linear order, and that the original type IIS restriction sites are lost after digestion, so the ligated product will not be subject to re-digestion (Figure 13-A).
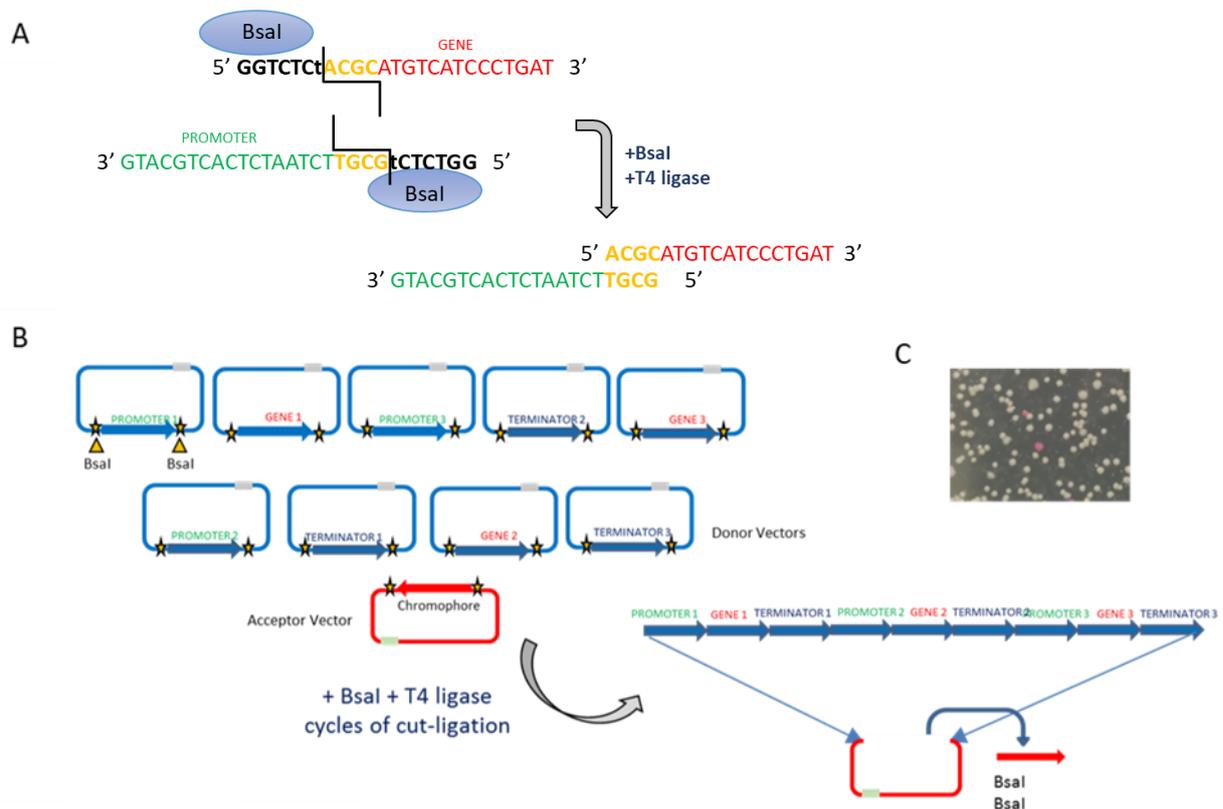
Figure 13. Golden Gate strategy.

A-Shows the digestion by BsaI and the compatible overhangs that are generated allowing the ligation between two fragments, here a promoter and a gene **B-** Schematic representation of the one-pot single-step modular cloning strategy. Donor vectors bearing building blocks with appropriate BsaI-site (Star) and pre-designed overhangs are mix together with an accepting vector bearing a red chromophore gene (RFP), BsaI restriction enzyme and T4 ligase. The custom-designed 4nt overhangs allow the oriented assembly of building blocks inside the destination vector. **C-** Image of *E. coli* transformed with the cloning product. The RFP reporter gene allows a rapid visual screening of positive clones.

The GG system can be divided into two main steps. First, the construction of a library containing the interchangeable building blocks ready to be used. Second, the assembly reaction itself, were the parts are connected to each other in a specific order to build up the desired DNA sequence.

In the GG system developed for *Y. lipolytica*, the DNA modules are assembled based on a scaffold of predesigned thirteen unique 4-nucleotide (nt) overhangs sequences that allows the oriented assembly of up to three TUs (each bearing a promoter, a gene, and a terminator), a selective marker, target sequences for genome integration, and a destination vector backbone. The method is based on BsaI restriction enzyme and was developed in a way that permit the assembly of 1, 2 or 3 TUs (Figure 14). Furthermore, the acceptor vector bears a red chromophore gene (RFP) between the BsaI restriction sites. When DNA fragments are assembled with the acceptor vector, the chro-
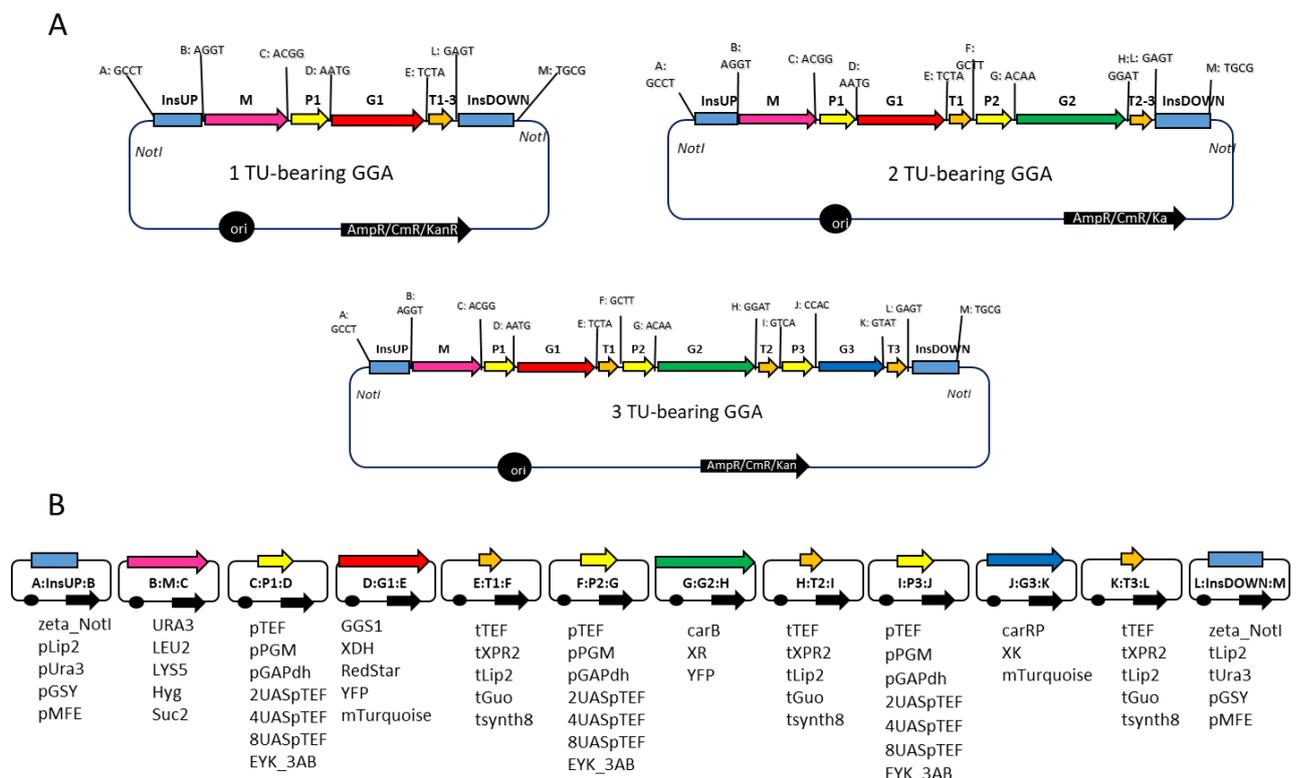
Figure 14. Golden Gate scaffold and library developed for *Y. lipolytica*, allowing the assembly of up to thirteen DNA blocks.

A-Schematic representation of the GG scaffolds that allow the assembly of 1, 2 o 3 TU. **B**- Example of building blocks that are available for each GG position. TU: transcription unit, composed of promoter (P), gene as ORF (G), and terminator (T). M: selection marker. InsUP and InsDOWN: genome integration targeting sequences upstream and downstream the cassette respectively. Letters A to M are assigned to each pre-designed 4nt overhang flanking the building blocks. Sequences of the overhangs are provided next to the respective position GGA: Golden Gate assembly. Figure adapted from Celinska et al., 2017

mophore is removed and replaced by the DNA fragments. Thus, the loss of the RFP reporter gene allows a rapid visual screening of positive *E. coli* clones and speeds up the assembly validation (Figure 13-B and C) (Celińska et al., 2017; Larroude et al., 2019).

During the present work a detailed protocol was set-up, a large library of DNA parts was constructed and characterized, and proof of concept constructions were carried out to evidence the increased efficiency on strains construction. These works were published as scientific articles, and a book chapter that is now under review, and will be presented hereafter.

The first step was to design and construct DNA parts, with the consequent construction of a library with reusable building blocks (Figure 14-B). Thus, DNA parts that are widely used across different constructions, such as selection markers, promoters, terminators, genome-insertion sequences and even reporter genes, are constructed only once and stored in a dedicated *E. coli* library for easy access and reuse. This library is permanently expanding as new parts are constantly being developed depending on the needs of each project. After the construction of the first library elements, parts were tested and characterised in *Y. lipolytica*. As a result, a characterised toolkit, composed of 64 building blocks, is available for the scientific community at Addgene https://www.addgene.org/ (nos. 120730-120793). This toolkit for DNA assembly dedicated to *Y. lipolytica* allows the entire construction of integrative expression cassettes, containing one, two or three genes.

The constructed kit contains three native promoters and six hybrids promoters from which three are inducible. As stated in section 1.3, the hybrid promoters are constructed by associating multiple UAS elements to a core promoter. Here, their expression level was evaluated using the fluorescent RedStarII protein as reporter gene, using our standard one TU-GG cloning strategy. The hybrid promoters showed higher expression than native promoters and the expression level was positively correlated with the number of UAS. In total there are nine promoters, that display weak to very strong expression strength, for the three TUs. Together with promoters, terminators are major components of gene expression systems. The toolkit includes five different terminators, Tlip2, Ttef, Txpr2 TsynthGuo and TSynth8, which impact on the expression level was analysed using pTEF as promoter and RedStarII as reporter gene, using the one TU GG cloning strategy. The results showed that the construction containing Tlip2 resulted in the greatest expression while the one with Ttef resulted in the lowest expression. As a result, combining promoters and terminators, this toolkit allows fine-tune expression of genes.

Furthermore, using various combinations of promoters and terminators limits the occurrence of homologous sequences inside the expression cassette and, thus, reduces the risk of recombination and the loss of fragments.

Another important part of the expression cassettes are selection markers. The markers available in the kit include three auxotrophic markers (URA3, LEU2 and LYS5), two antibiotic resistance markers (Hygromycin and nourseothricin resistance genes) and a metabolic marker (SUC2). These markers are designed to be flanked by Lox sequences allowing their easy excision when expressing the Cre recombinase, thus markers are recycled and can be used in a future transformations (Fickers et al., 2003). Nonetheless, the large panel of markers included in the kit allows numerous successive transformations without the need of recycling markers. In addition, they allow performing transformations in a wide range of background strains including wild types.

The system also includes several sequences to flank the expression cassette, allowing the integration at either random or specific locus. As introduced in sections 1.2.1 and 1.3, *Y. lipolytica* needs homologous recombination sequences of 1kb long, for DNA integration at specific locus. This HR allows the insertion of an expression cassette at a specific locus while simultaneously deleting a target gene of particular interest. The sequences for site specific integration that are included in the kit target genes involved in the lipid metabolism (*LIP2*, *GSY* and *MFE*) which is one of the most explored applications of this oleaginous yeast. However, new sequences can be easily constructed to target other genes. For random integration, the widely used zeta sequences were included.

As part of the validation process of the toolkit, the expression level of proteins was evaluated in a 1 and 3 TU assembly approach using three different fluorescent proteins (RedStarII, YFP and mTurquoise). Proteins within an assembly are well expressed, however, we observed that the expression level can be affected by the position on the assembly.

In addition, three genes allowing the use of xylose as carbon source (ylXK, ylXR and ylXDH) by *Y. lipolytica* were overexpressed as a full pathway integration proof of concept. Even though endogenous genes, they have to be overexpressed in order to allow the use of xylose. The overall process, from cloning to phenotype screening, took

less than 10 days, which is much faster than carrying out sequential plasmid transformation, estimated to be at best one week per gene. This shows the capability of fast metabolic rewiring of *Y. lipolytica* using the GG technique.

The complete results are presented in the following article published in Microbial Biotechnology (Larroude et al., 2019) and supplementary data of this work containing protocols and the list of strains and primers used is available on the annexes part of this manuscript (6.1.1).

microbial biotechnology

A modular Golden Gate toolkit for *Yarrowia lipolytica* synthetic biology

Macarena Larroude,[1] Young-Kyoung Park,[1]
Paul Soudier,[1] Monika Kubiak,[2] Jean-Marc Nicaud[1]
and Tristan Rossignol[1],* (iD)

[1]*Micalis Institute, AgroParisTech, INRA, Université Paris-Saclay, 78350, Jouy-en-Josas, France .*
[2]*Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, ul. Wojska Polskiego 48, 60-627, Poznan, Poland .*

Summary

The oleaginous yeast *Yarrowia lipolytica* is an established host for the bio-based production of valuable compounds and an organism for which many genetic tools have been developed. However, to properly engineer *Y. lipolytica* and take full advantage of its potential, we need efficient, versatile, standardized and modular cloning tools. Here, we present a new modular Golden Gate toolkit for the one-step assembly of three transcription units that includes a selective marker and sequences for genome integration. Perfectly suited to a combinatorial approach, it contains nine different validated promoters, including inducible promoters, which allows expression to be fine-tuned. Moreover, this toolbox incorporates six different markers (three auxotrophic markers, two antibiotic-resistance markers and one metabolic marker), which allows the fast sequential construction and transformation of multiple elements. In total, the toolbox contains 64 bricks, and it has been validated and characterized using three different fluorescent reporter proteins. Additionally, it was successfully used to assemble and integrate a three-gene pathway allowing xylose utilization by *Y. lipolytica*. This toolbox provides a powerful new tool for rapidly engineering *Y. lipolytica* strains and is available to the community through Addgene.

Introduction

Yarrowia lipolytica is the most well-developed and well-researched yeast in the domain of oleochemical production (Beopoulos *et al.*, 2009; Ledesma-Amaro and Nicaud, 2016; Lazar *et al.*, 2018). It is considered to be a GRAS organism (Groenewald *et al.*, 2013) and has an established history within the biotechnology industry. It has been intensively used for various applications, ranging from biofuel to vaccine production (Madzak, 2018). However, when the goal is to produce a compound at the industrial scale, multiple rounds of metabolic engineering are usually needed. For example, for *Y. lipolytica* to produce omega-3 eicosapentaenoic acid at industrial levels, it has been necessary to integrate up to 30 copies of nine different genes and carry out one deletion (Xue *et al.*, 2013). Similarly, laboratory-scale production of ricinoleic acid has required the overexpression of three genes and the deletion of up to 10 genes, mainly due to the redundancy of genes involved in lipid metabolism in oleaginous microorganisms (Beopoulos *et al.*, 2014). Therefore, when seeking to re-engineer such specialized microorganisms, it can be extremely tedious, time-consuming and cost-ineffective to work with the existing genetic background. It requires massive efforts to integrate an equivalent series of modifications into a new wild-type strain that presents specific or more appropriate traits with a view to creating new producer or chassis strains. Moreover, strain reconstruction requires multiple steps that sometimes lead to a final construct that does not display the expected phenotype or production yield because the transformation process results in the accumulation of potential trade-offs resulting from random insertions for example.

Moreover, to obtain an optimized producer strain using the design–build–test–learn cycle, it is necessary to perform large-scale screening using the combinatorial expression of pathway components and to fine-tune control of gene expression. Rapid, efficient and combinatorial cloning tools are needed for such approaches.

For these reasons, researchers are developing versatile, standardized and modular tools for carrying out genetic engineering and genome editing in *Y. lipolytica*, as such tools have recently become available in the model yeast *Saccharomyces cerevisiae* (Lee *et al.*, 2015). Two key tools have recently been released for *Y. lipolytica*: EasyCloneYALI, which is based on an

alternative USER cloning approach (Holkenbrink *et al.*, 2018), and the YaliBricks system (Wong *et al.*, 2017), which utilizes the BioBricks standard and four compatible restriction enzyme sites. Here, we present the development and release of a new modular toolkit dedicated to *Y. lipolytica* that is based on the popular Golden Gate (GG) strategy (Engler *et al.*, 2009). It allows the one-step assembly of up to three transcription units (TUs) together with a selective marker and sequences for random or targeted genome integration. The toolkit has been designed for optimum modularity and flexibility: it makes available nine promoters of varying strengths, including inducible promoters that are perfectly suited to industrial applications. All the promoters are available at each TU position. The toolkit also includes five terminators, six selection markers (three auxotrophic markers, two antibiotic-resistance markers and one metabolic marker) and four pairs of bricks for random or targeted genome integration. A detailed description is provided of GG brick design and assembly (see supplementary protocol S4). All the bricks are available to the community through Addgene. Our toolkit is a powerful new tool for rapidly engineering *Y. lipolytica* strains.

Results

Brick building

The GG strategy used in our *Y. lipolytica* toolkit utilizes the BsaI Type IIS restriction enzyme. Thirteen unique overhang sequences were designed that allow the oriented assembly of up to three TUs together with a selection marker, upstream and downstream sequences for random or targeted genome insertion, and the *Escherichia coli* replicative backbone. Each TU is composed of a promoter, a gene of interest and a terminator. An assembled vector is depicted in Fig. 1; the letters indicate the four nucleotide overhang sequences (described in Fig. S1.)

Promoters. To fine-tune and allow combinatorial expression of each TU in this assembly system, promoters with a broad range of strengths must be implemented. Therefore, six promoters that displayed weak to very strong expression strength were constructed. There were three native promoters: the widely used p*TEF* promoter and two other lower-strength promoters, p*PGM* and p*GAP* (Larroude *et al.*, 2018).
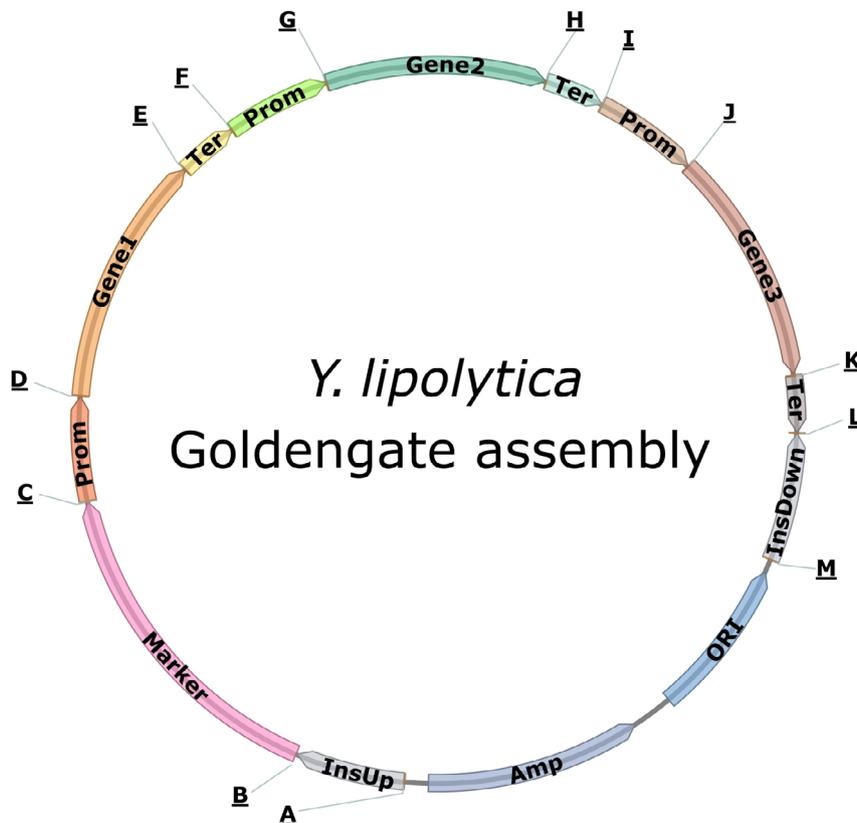
Fig. 1. Schematic drawing of the 3-TU GG assembly with letters indicating the 4-nt overhang flanking each GG brick. The sequences of each 4-nt overhang are available in Fig. S1. InsUp and InsDown: sequences for insertion in the genome; Prom: promoter; Ter: terminator.

There were also three synthetic hybrid promoters, which were composed of a core promoter preceded by two to eight repeated UAS sequences (Dulermo et al., 2017). In addition, we included three hybrid promoters that are inducible by erythritol. They are based on the recently developed pEYK1 promoter and incorporate three to five repeated UAS sequences (Trassaert et al., 2017; Park et al., 2019). Consequently, in total, there are nine promoters available for the three TUs.

Figure 2 shows the expression levels associated with the nine promoters when RedStarII was employed as the reporter gene in position 1 (first TU); glucose was the carbon source for the constitutive promoters (Fig. 2A), and glucose and erythritol were the carbon sources for the inducible promoters (Fig. 2B). Under these conditions, expression levels were lower with pPGM and pGAP than with pTEF. In contrast, expression levels were relatively higher with all the synthetic promoters, and the promoters with four and eight UASs performed better than the one with two UASs. The 8-UAS promoter and the 4-UAS promoter resulted in similar expression levels and thus displayed similar performance in this condition. Either we achieved maximum expression under our conditions with 4UAS pTEF and 8UAS pTEF or the latter was not the best suited promoter for the RedStarII under these conditions (Dulermo et al., 2017). For the inducible promoters, expression level was positively correlated with UAS1-eyk1 copy number when the medium contained erythritol; in contrast, expression level remained very low when the medium contained only glucose, as previously reported (Park et al., 2019).

Terminators. Terminators are major components of expression systems because they are responsible for terminating transcription and because they play a role in mRNA half-life. In our toolkit, we have included five

different terminators (Celinska et al., 2017) (Fig. S1). To characterize their impact on expression level, all the terminators were tested in position 1 (first TU) using the pTEF promoter and the RedStarII reporter gene. We found that each one allowed the correct expression of the reporter gene, but there were differences in expression strength: TLip2 resulted in the greatest expression, and TTef resulted in the lowest expression (Fig. 3). The role of terminators in mediating expression levels has been extensively described in *Saccharomyces cerevisiae* (Curran et al., 2015; Redden et al., 2015) and somewhat characterized in *Y. lipolytica* (Curran et al., 2015). The terminators T1Guo and TSynth8 resulted in similar levels of expression, as reported previously (Curran et al., 2015). In this study, however, expression levels were lower with TTef than with T1Guo and TSynth8, while the opposite was true in Curran et al. (2015). This discrepancy can be explained by the fact that the latter study used a different expression cassette, one that bore the HrGFP reporter
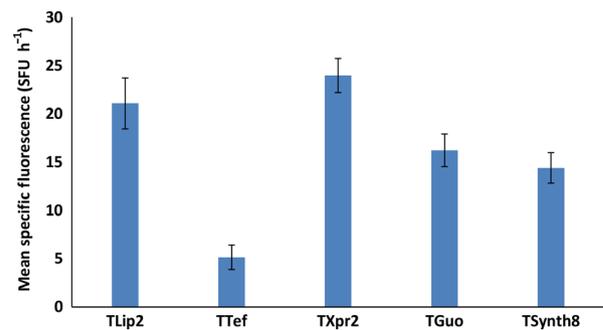
Fig. 3. Impact of different terminators on gene expression levels using *RedStarII* as a reporter gene under the pTEF promoter. The bars correspond to mean specific red fluorescence (position 1). The values correspond to the mean of five to seven independent clones that randomly integrated each construction. JMY1212 strain was used. Error bars represent standard deviations.
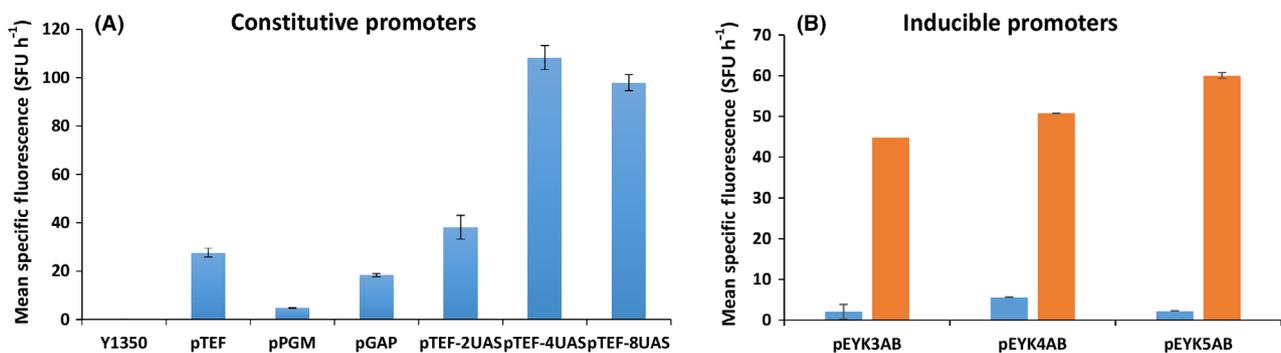
Fig. 2. Expression associated with the nine promoters using *RedStarII* as a reporter gene in position 1 (first TU). A. Constitutive promoters, using glucose as the carbon source. B. Inducible promoters, using glucose (blue bars) and erythritol (orange bars) as the carbon sources. Promoters were tested on the strain JMY1212, and a construct-free strain (JMY1350; Table S1) was the control. The values correspond to the mean of two independent clones that randomly integrated each construct. Error bars represent standard deviations.

gene and a hybrid promoter. More importantly, the terminator sequences were longer than the ones used here (253 base pairs versus 66 base pairs). As a result, this toolkit allows expression to be fine-tuned using various combinations of terminators and promoters. Furthermore, this approach helps limit the occurrence of homologous sequences inside the vector and thus reduces the risk of recombination and the loss of part of the constructs. To allow the assembly of just one or two TUs, we have also provided T*Lip2* terminators containing the BsaI overhang sequences E and L and terminators containing the BsaI overhang sequences H and L; the result is that there is no need to change the overhang sequences of the brick genes (Fig. S1).

Markers. The core principle of the GG strategy is modularity and versatility. Consequently, our set of bricks includes a panel of markers that meet the requirements of most *Y. lipolytica* genetic backgrounds. We implemented three auxotrophic markers, *URA3*, *LEU2* and *LYS5*, which are the most popular auxotrophic markers used with laboratory strains. Two markers for antibiotic resistance (against hygromycin and nourseothricin respectively) are also available; they can be used with auxotrophic strains as well as with wild-type strains. In addition, the toolkit includes a metabolic marker, the invertase from *S. cerevisiae*, which allows growth on sucrose (Lazar *et al.*, 2013). This large panel of markers makes it possible to carry out numerous successive transformations without the need to recycle markers. Moreover, the latter three markers can be used to modify wild-type strains. All the markers were successfully used in the assembly process with a RedStarII expression cassette (Table S1) and were transformed into an appropriate background for validation (data not shown).

Integration sites. To extend the genetic engineering capabilities of our system, we have included several flanking sequences that allow integration at either a random locus or a specific locus. ZETA sequences are widely used and allow random integration; however, they also allow targeted integration into the zeta-docking platform when it is present in strains. Such strains are employed in various bioprocesses (e.g. JMY1212; Bordes *et al.*, 2007) or are used to carry out the large-scale screening of homologous or heterologous gene overexpression (e.g. JMY2566; Leplat *et al.*, 2015; Beneyton *et al.*, 2017). When using ZETA sequences, the presence of the docking platform strongly favours integration, with success rates of up to 84% (Bordes *et al.*, 2007).

In addition, there is an industrial interest in deleting certain genes, namely *LIP2*, *GSY1* or *MFE*. Consequently, their promoter and terminator regions were added to the list of insertion sites. This approach allows an expression cassette of interest to be integrated while simultaneously deleting a gene or pathway that interferes with biotechnological applications in the domain of lipid metabolism. Moreover, we choose targets for which phenotype screening is easier: the halo is reduced on plates containing tributyrin in the case of *lip2*Δ strains (Pignede *et al.*, 2000); cells appear less brown when exposed to Lugol's solution in the case of *gsy1*Δ strains (Bhutada *et al.*, 2017); and cells cannot grow on media containing lipids as the only carbon source in the case of *mfe*Δ strains. We evaluated the integration of a RedStarII expression cassette under a p*TEF* promoter flanked by LIP2 or GSY1 insertion site sequences. For LIP2, among the 37 transformants that showed red fluorescence, five displayed the Δ*lip2* phenotype on a tributyrin plate (see the example in Fig. 4A) and reflected proper integration at the lip2 locus via homologous recombination, which corresponds to a 14% integration rate. For GSY1, among the 24 transformants that showed red fluorescence, 11 displayed the Δ*gsy1* phenotype after Lugol's iodine staining (see the example in Fig. 4B) and reflected proper integration at the gsy1 locus via homologous recombination, which corresponds to a 45% integration rate. The integration rate at the lip2 locus was unsurprising because *Y. lipolytica* has a low homologous recombination efficiency. In contrast, we have always obtained higher recombination rates for the gsy1 locus (data not shown). The RedStarII fluorescence level is also much higher when expressed at the gsy1 locus compared with the lip2 locus (Fig. 4C). This latter shows similar level to random integration.

Users can easily build upon this list, by adding flanking sequences of their own that target specific DNA regions; we provide a detailed protocol in the Supplementary material.

Destination vector. The two destination vector backbones provided in this GG toolkit contain the red fluorescence protein (RFP) chromophore, which acts as a colour-based visual marker for negative cloning in *E. coli*, as described elsewhere (Celinska *et al.*, 2017). The vector on GGE029 only contains the ampicillin resistance, the ColE1 region for selection and propagation in *E. coli* and the RFP flanked with the BsaI sites required (A and M) to assemble the expression cassettes. The vector on GGE114 is a preassembled destination vector that, in addition to the bacterial part, contains popular bricks – ZETA sequences in the place of InsUp and InsDown fragments and the *URA3* marker – with a view to reducing the number of fragments to assembly when employing this combination, which is the most common (Park *et al.*, 2019). In this case, the RFP is between the *URA3* marker and the ZETA down. In the
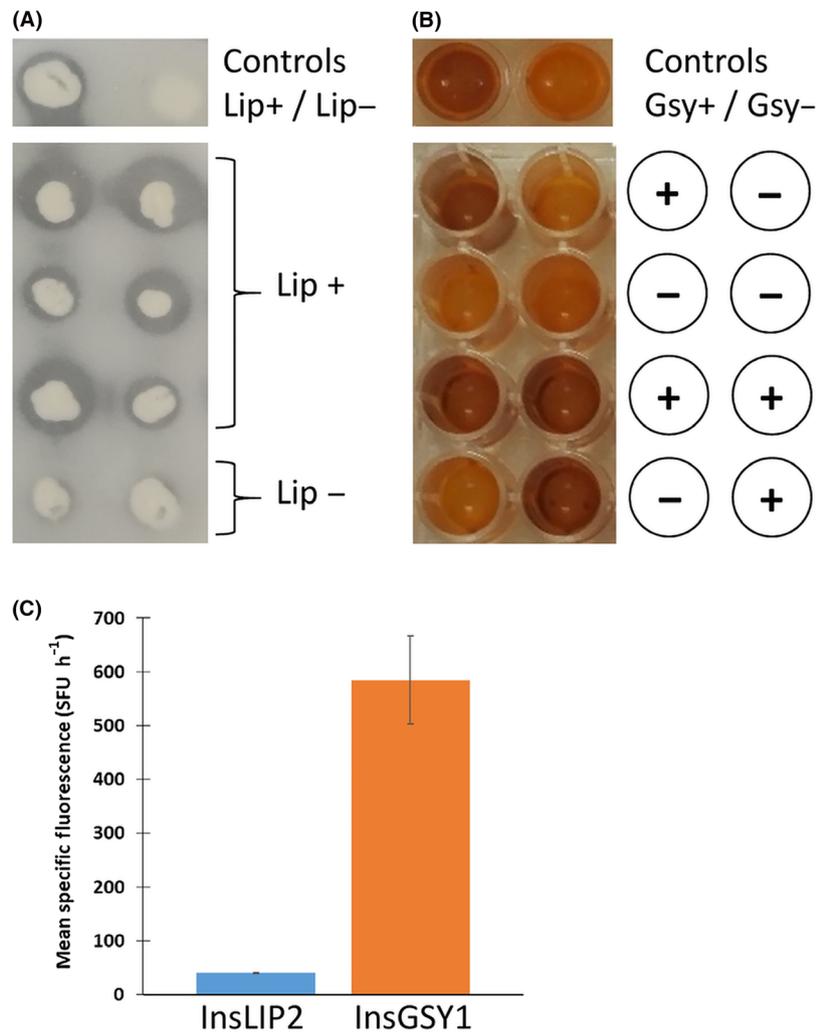
Fig. 4. Examples of successful integration at the locus sites. A. Detection of lipase activity on a tributyrin plate. When the colonies were surrounded by a clear zone, they were capable of lipase production. The GGA was used to transform a JMY195 strain, and the results were compared with those for the wild type, W29, and the lipase defective strain, JMY1212. B. Detection of glycogen synthase activity using Lugol's solution. When yellow wells were present, it indicated a defect in glycogen synthase, which is a characteristic of the Δgsy1 mutant strain (Bhutada *et al.*, 2017). In this case, the JMY1212 strain was used to test the integration at gsy locus. These results were compared with those for the wild type, W29. C. Impact of integration site on gene expression levels using *RedStarII* as a reporter gene. The bars correspond to mean specific red fluorescence. The values correspond to the mean of four independent clones that have a correct locus integration. Error bars represent standard deviations.

presence of BsaI, the RFP is released and the TU can be added in that place. Thus, both destination vectors can be used for assembling 1, 2 or 3 TU.

Protein expression

The capacity and efficacy of our system to express three TUs on a same construct has been validated previously (Celinska *et al.*, 2017; Larroude *et al.*, 2018). However, the expression levels of the three TUs have not been evaluated. Here, we used three fluorescent proteins as reporters to validate and quantify TU expression levels.

Expression of the three fluorescent reporter proteins. First, we evaluated the expression and detection of the three fluorescent proteins – RedStarII, YFP and mTurquoise – in position 1 in the *Y. lipolytica* GG system using p*TEF* and T*Lip2*. The fluorescence observed at these three wavelengths shows that the three proteins were correctly expressed (Fig. 5).

To evaluate expression levels at the different positions associated with the three TUs, the three proteins were assembled together in the same 3-TU vector and using the same promoter (p*TEF*) and terminator (T*Lip2*) as in the above, single-position experiment. RedStarII
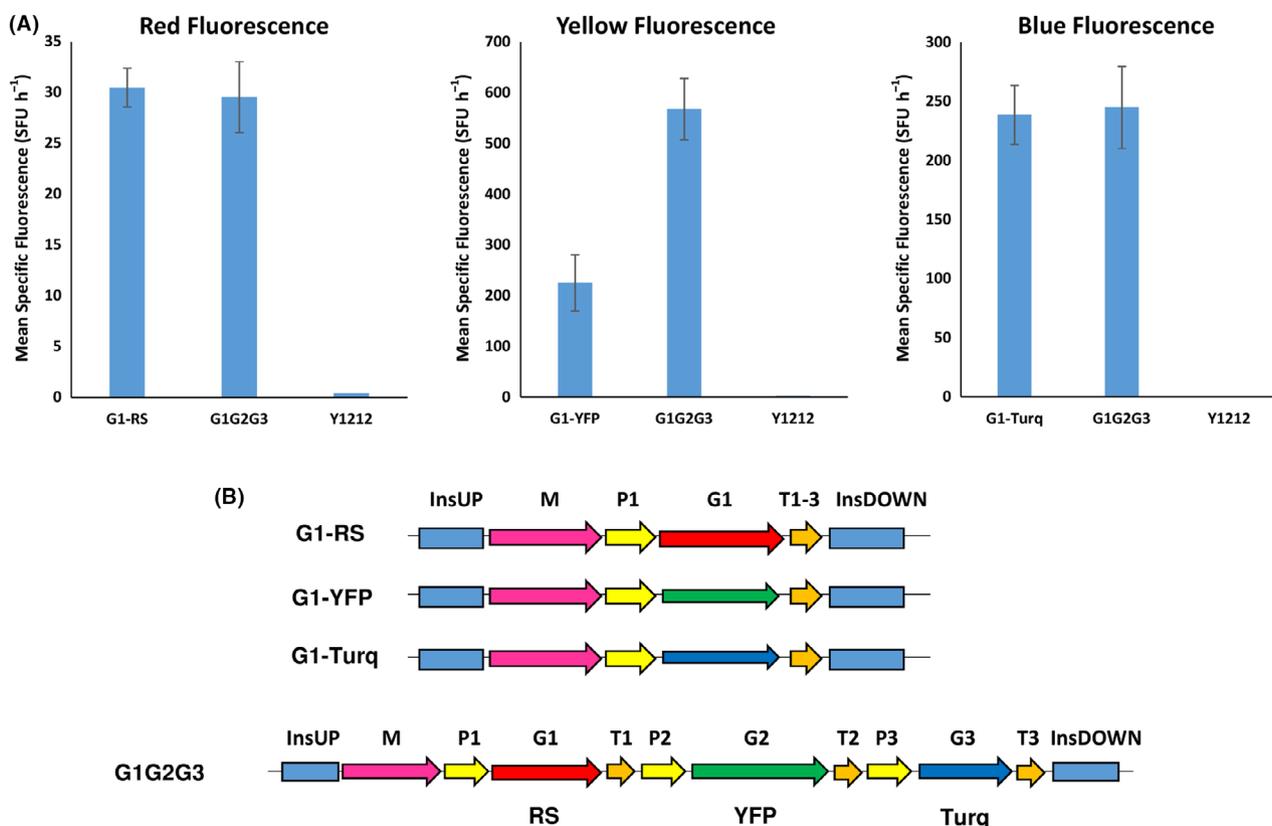
Fig. 5. A. Mean specific fluorescence of strains expressing a single fluorescent protein – RedStarII, YFP or mTurquoise – in position 1 of a 1-TU GG vector or of strains expressing all three fluorescent proteins in a 3-TU GG vector. In G1G2G3, *RedStarII* is in position 1, *YFP* is in position 2, and *mTurquoise* is in position 3. JMY1212 was the control strain. The values correspond to the average for 8–10 independent clones that randomly integrated each construct. Error bars represent standard deviations. B. Schematic of the construct used, all of them were transformed into JMY1212 strain.

occurred at position 1, YFP occurred at position 2, and mTurquoise occurred at position 3. The three proteins were correctly expressed (Fig. 5). The expression levels of RedStarII and mTurquoise were similar to those obtained in the single-position experiment; in contrast, the expression level of YFP was higher. We then switched around the positions of the fluorescent proteins. RedStarII and mTurquoise expression was unaffected by position; YFP behaved differently but only when it was in position 2 (data not shown). This finding reveals that, for some proteins, expression levels could be TU position-dependent.

Assembly and expression of the xylose utilization pathway. Yarrowia lipolytica cannot naturally use xylose even though it has genes that code for the utilization pathway. Xylose is a major component of lignocellulosic material, and microbial cell factories must be able to exploit this compound as a carbon source. To demonstrate the utility of the GG tool in this context, we assembled an overexpressed xylose utilization pathway that was based on three genes: one for xylitol

dehydrogenase *XDH* (YALI0E12463g), one for xylose reductase *XR* (YALI0D07634g) and one for xylulokinase *XK* (YALIF10923g). This pathway allows *Y. lipolytica* to grow using xylose as its sole carbon source (Niehus et al., 2018). The three genes were assembled in a single plasmid (Fig. 6A) using the protocol we describe in the supplementary material. The strain Y1212 was then transformed with this construct, and its ability to grow on xylose was evaluated. We found that 79% of the transformants correctly expressed the xylose cassette (Fig. 6B). The overall process, from cloning to phenotype screening, took less than 10 days, which is much faster than carrying out sequential plasmid transformation, in which it is impossible to verify phenotypes before the three genes have been co-expressed.

Discussion

Here, we present a Golden Gate toolkit for *Y. lipolytica* that can be used to rapidly assemble multigene pathways. It is a powerful new tool for one-step strain
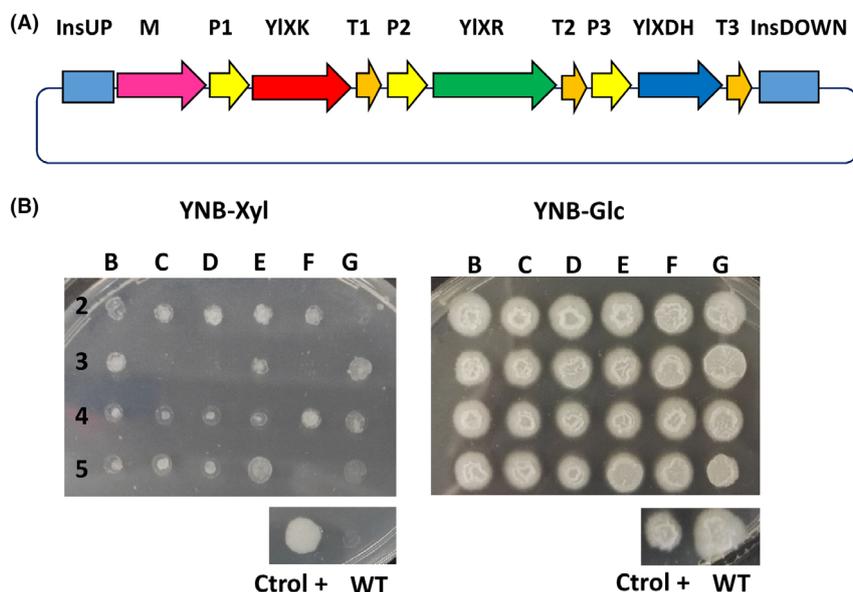
Fig. 6. Synthetic pathway assembly using the Golden Gate toolkit and the resulting expression in *Y. lipolytica*. A. Schematic representation of a three-gene assembly, composed of *Y. lipolytica* xylulokinase (ylXK), *Y. lipolytica* xylitol dehydrogenase (ylXDH) and *Y. lipolytica* xylose reductase (ylXR) genes, that allows *Y. lipolytica* to grow on xylose. B. *Y. lipolytica* clones after transformation with the xylose cassette that were grown in a plate containing only xylose as the carbon source; clones were grown in a plate containing glucose as the carbon source for the control. The xylose cassette was correctly expressed in 79% of JMY1212 transformants. WT: wild type. Xyl: xylose. Glc: glucose. Ctrol+: positive control, which was the *Y. lipolytica* strain expressing the three genes developed via the standard cloning method and whose xylose utilization was verified (Ledesma-Amaro *et al.*, 2016).

engineering. A colour-based reporter system is included in the backbone vectors, which reduces the possible number of false-positive clones during the selection process and thus speeds up assembly validation (see supplementary protocol S4). The toolkit is available to the research community through Addgene, under the plasmid ID numbers 120730-120793, and will be regularly updated. At present, it contains 64 GG bricks: 27 promoter bricks; 14 terminator bricks; 6 markers; 10 genome-insertion sequences corresponding to three different insertion loci and one for random integration; five gene bricks coding for fluorescent proteins that can serve as validation genes in positions 1, 2 or 3; and two destination vectors (see Fig. S1). The versatility and standardization of these bricks means that the toolkit can easily be expanded by other research teams, using our detailed protocol (see the supplementary material).

The BsaI sites used in our system are not compatible with the MoClo toolkit for *S. cerevisiae* (Lee *et al.*, 2015), but expression of *S. cerevisiae* genes or optimized for *S. cerevisiae* is sometimes not functional in *Y. lipolytica* as they are genetically distant. Consequently, tools developed for *S. cerevisiae* cannot be systematically used in *Y. lipolytica*. For example, when we used the mTurquoise fluorescent protein from the *S. cerevisiae* MoClo toolkit, it was not functional in our system (data not shown). In contrast, when we used the version of mTurquoise optimized for *Y. lipolytica*, the

protein was expressed correctly (Fig. 5). Likewise, we attempted to use the p*TEF* promoter from the *S. cerevisiae* MoClo toolkit when expressing RedStarII, but it was non-functional in *Y. lipolytica* (data not shown). These results highlight the risks that bricks will be incompatible between these two hosts.

Other multigene assembly systems dedicated to *Y. lipolytica* have been described recently. The YaliBricks system (Wong *et al.*, 2017) uses the BioBricks standard and four compatible restriction enzyme sites for modular pathway assembly. The tool contains 12 natural promoters, of which p*TEF* is the strongest. In this system, the use of a fluorescent reporter was unsuccessful because the signal was too weak, and luminescence had to be used instead (Wong *et al.*, 2017). Our system provides a larger range of expression levels via the deployment of synthetic promoters, which are well described here (Fig. 2). An additional advantage of synthetic promoters is that they are probably much less susceptible to as-yet-unknown regulatory mechanisms, which is not the case for endogenous promoters of genes from the lipogenic pathway used in Wong *et al.* (2017). This problem may be especially important in oleaginous organisms. Moreover, our system seems less susceptible to the transcription inhibition that can occur when multigene constructs are under the control of the same promoter, a phenomenon observed by Wong *et al.* (2017). Another point of contrast is that, while the

YaliBricks system has been used to successfully assemble five genes from the violacein pathway, it relies on sequential gene assembly. In our GG system, the use of a single restriction site means that assembly is performed in a single 'step'. Indeed, with a one-pot reaction, we are able to assemble three TUs, a marker and various integration site sequences, which correspond to the assembly of twelve fragments in a destination vector.

Very recently, Holkenbrink *et al*. (2018) developed a toolkit based on an alternative USER cloning approach, which allows the integration of one to two TUs at intergenic regions called EasyCloneYALI. The system largely utilizes a ku70 mutant background and a CRISPR/Cas9 strategy, and it carries out improved homologous recombination at specific non-coding loci. However, the ku70 mutant has a low transformation rate (Verbeke *et al*., 2013), which might make it less suitable for the engineering of multiple and/or successive genes. Our system can be used with any genetic background, including wild-type strains. The EasyCloneYALI toolkit contains 14 different promoters. However, only eight of them functionally expressed a fluorescent protein. The nine promoters provided in our GG toolkit have been tested and validated, and their expression capacity has been characterized. Very recently, Egermeier *et al*. (2019) published a GG method for *Y. lipolytica* that is based on the GoldenMOCS system. Their method systemically requires several cloning steps, which make it fundamentally different from ours. Essentially, Egermeier *et al*. have used it to produce a CRISPR/Cas9 plasmid, as well as an expression vector containing one TU, which was based on a replicative plasmid. At this stage, it is much less modular and expandable than our GG system when it comes to the expression of multigene pathways.

Here, we engineered a xylose utilization pathway to highlight how our GG toolkit can greatly accelerate the build–test step of the classical design–build–test–learn cycle in synthetic biology. It is particularly important to be able to construct multi-TU plasmids because single genes may not yield the desired phenotype and the functionality of a single gene cannot be confirmed before successive transformations are carried out. The example presented here demonstrates that top-performing transformants can be screened in one step. Previously, we used a smaller version of this toolkit to illustrate the assembly of an entire pathway, notably the carotenoid pathway (Celinska *et al*., 2017; Celińska *et al*., 2018; Larroude *et al*., 2018), and the construction of a set of expression cassettes containing multiple secretion signal sequences, which allows the optimization of protein secretion (Celińska *et al*., 2018; Soudier *et al*., 2019). Here, we have greatly expanded the capacity of our system, and we now provide strong as well as inducible promoters. The toolkit also contains a large panel of promoters and terminators, which allows for high-throughput screening. More specifically, combinatorial TUs can be randomly assembled in donor vectors, generating a pool that can be screened for the best combination. Here, we have limited ourselves to providing bricks for a subset of lipid metabolism genes that could be disrupted. Indeed, we wish to underscore that our toolkit could be easily employed as a 'consolidated' approach, which carries out pathway integration and gene deletion simultaneously. However, in theory, any integration site sequences could be used. The goal is for users to be able to delete the targets of their choice, and our GG toolkit allows them to quickly and easily design and assemble cassettes that function simultaneously in disruption and overexpression.

Experimental procedures

Strains and media

The *Y. lipolytica* strains constructed and used in this study are listed in Table S1, as are the *E. coli* strains that hosted the GG-assembled vectors.

Escherichia coli strain DH5α was used for cloning and plasmid propagation. The transformation of chemically competent *E. coli* cells was performed using a heat shock protocol. Cells were grown at 37°C with constant shaking on 5 ml of LB medium (10 g l^{-1} tryptone, 5 g l^{-1} yeast extract and 10 g l^{-1} NaCl); ampicillin (100 μg ml^{-1}) or kanamycin (50 μg ml^{-1}) was added for plasmid selection.

Yarrowia lipolytica strain JMY1212 strain was used in this study. The YNBD minimal media contained 10 g l^{-1} glucose (Sigma-Aldrich, Saint-Quentin Fallavier, France), 1.7 g l^{-1} yeast nitrogen base (YNBww; Difco, Paris, France), 5.0 g l^{-1} NH4Cl and 50 mM phosphate buffer (pH 6.8). Glucose was replaced with erythritol (10 g l^{-1}) in the induction experiments or with xylose (2% wt/vol) for the purposes of verifying the xyl+ phenotypes. When necessary, the YNB medium was supplemented with uracil (0.1 g l^{-1}) or leucine (0.1 g l^{-1}). Solid media for *E. coli* and *Y. lipolytica* were prepared by adding 15 g l^{-1} agar (Invitrogen, Saint-Aubin, France) to liquid media.

Building the brick plasmids

Primers carrying predesigned 4-nt overhangs and externally located BsaI recognition sites were created and used to amplify the bricks, which were cloned in donor vectors (Zero Blunt® TOPO® PCR Cloning Kit, Thermo Fisher, UK) unless otherwise stated. They were then transformed into *E. coli* (see the Data S1). PCR amplifications were performed using Q5 high-fidelity DNA polymerase (NEB) or GoTaq DNA polymerase (Promega). The native sequences used as building bricks were amplified from

the genome of *Y. lipolytica* W29. When needed, PCR fragments were purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). The mTurquoise gene (Lee *et al.*, 2015) was codon optimized for *Y. lipolytica* using COOL online software (Chin *et al.*, 2014). The *Streptomyces noursei nat1* gene (conferring nourseothri-cin resistance), the *E. coli hph* gene (conferring hygromy-cin resistance) (Tsakraklides *et al.*, 2018), the *RedStarII* gene and the *YFP* gene were adapted for the GG system by eliminating internal BsaI recognition sites and adding the corresponding overhangs. They were synthesized by Twist Bioscience and then cloned into a TOPO vector. The genes *nat1* and *hph* were already codon optimized for *Y. lipolytica*, and their corresponding bricks contained p*TEF* and T*Lip2* for expression in *Y. lipolytica*. T*Synth8* and T*Guo* (Curran *et al.*, 2015) were synthetized by annealing single-strand sequences, and they were then cloned into a TOPO vector.

   All donor vectors were verified by restriction profile anal-ysis (BsaI) and by sequencing. All restriction enzymes were purchased from New England Biolabs (NEB), and sequencing was carried out by Eurofins Genomics. The primer sequences used in this study for DNA module con-struction can be found in the Data S1, and the complete list of plasmids is available in Table S1. The sequences of all the building bricks can be found in File S1. The genes in the xylose pathway were synthesized and adapted for the GG system by eliminating internal BsaI recognition sites and adding the corresponding overhangs; they were then cloned in TOPO or pUC57 vectors, giving rise to GGV pUC57 G1 *XDHno*-BsaI, GGV pUC57 G2 *XRno*-BsaI and GGE 0097 TOPO G3 *XKno*-BsaI. The three genes were assembled using the GG protocol (see the Data S1), giving rise to plasmid named GGE0106 (ZUp-NotI_*URA*_P1*Tef_XDH*_T1*Lip2*_P2T*ef_XR*_T2*Lip2*_P3-*Tef_XK*_T3*Lip2*_ZDNotI).

GG cloning procedures

Plasmids from *E. coli* were extracted using the QIAprep Spin Miniprep Kit (Qiagen). All the reactions were per-formed according to the manufacturer's instructions. The plasmids containing the building bricks and the destination vector (flanked with the BsaI site and predesigned over-hangs) were added in equimolar quantities (50 pmoles of each DNA fragment to be assembled) to a one-pot reac-tion together with 5 U of BsaI (NEB), 200 U of T4 ligase (NEB), 2 μl of T4 DNA ligase buffer (NEB) and up to 20 μl of ddH$_2$O. The following thermal programme was used: [37°C for 5 min, 16°C for 2 min] × 60, 55°C for 5 min, 80°C for 5 min and 15°C ∞. Afterwards, 10 μl of the reac-tion mixture was used for *E. coli* transformation. White colonies were screened for correct GG assembly by col-ony PCR; plasmid isolation and restriction digestion with

NotI were then performed for verification. The detailed protocol is described in the supplementary material. Over-all, when 1 to 2 TUs were assembled (corresponding to 7–10 fragments), 50% of the *E. coli* clones showed the white phenotype, and 90% of them displayed correct assembly. When three TUs were assembled (corresponding to 13 fragments), 20% of the *E. coli* clones showed the white phenotype, and 80% of them displayed correct assembly.

   To improve assembly efficiency for the three TUs, the procedure was split into two parts: creation of preassem-bly constructs and assembly of the multigene construct. During preassembly, we carried out three separate reac-tions with four GG parts each (InsertionSiteUp-Marker-Promoter1-Gene1, Terminator1-Promoter2-Gene2-Termi-nator2 and Promoter3-Gene3-Terminator3-Inser-tionSIteDown). All the parts were present in equimolar quantities (50 pmoles). Each reaction also included 5 U of BsaI, 200 U of T4 ligase, 1 μl of T4 DNA ligase buffer (NEB) and up to 10 μl of ddH$_2$O. A short thermal pro-gramme was used: [37°C for 3 min, 16°C for 2 min] × 30, 55°C for 5 min, 80°C for 5 min and 15°C ∞. During the assembly of the multigene construct, the three previous reactions were mixed together in the same tube along with the destination vector (50 pmoles), 20 U of BsaI, 400 U of T4 ligase, 4 μl of T4 DNA ligase buffer (NEB) and up to 40 μl of ddH$_2$O. The following thermal programme was used: [37°C for 5 min, 16°C for 5 min] × 50, 55°C for 5 min, 80°C for 5 min and 15°C ∞. Subsequently, 15 μl of the GG reaction was trans-formed into *E. coli* and plated in a selective medium. White colonies were screened for correct GG assembly by colony PCR. Plasmid isolation and restriction diges-tion with NotI were performed for verification. When this method was employed, the rate of correct assembly was similar to that obtained for 1–2 TUs.

Construction of Y. lipolytica *strains*

Correct GG assemblies were subsequently linearized using the NotI restriction enzyme to allow the release of the expression cassette, and 10 μl was used to trans-form *Y. lipolytica* JMY1212, yielding prototrophic trans-formants. Transformation was performed using the lithium-acetate method (Le Dall *et al.*, 1994), and trans-formants were selected using YNB medium. To screen for antibiotic resistance, transformation reactions were plated on YPD containing hygromycin (200 μg ml^{-1}) or nourseothricin (500 μg ml^{-1}).

Growth and fluorescence analysis

Yarrowia lipolytica precultures were grown for 24 h in YNBD medium (supplemented with uracil when needed) in 96-well plates. Two μl was then transferred into 200 μl

of fresh medium in 96-well microplates (OD$_{600nm}$ of 0.1). YNB medium, supplemented with glucose (10 g l^{-1}, YNBD) or erythritol (10 g l^{-1}, YNBE), was used in the growth and fluorescence analysis (i.e. the choice depended on the promoters). The growth analysis was performed using a microtitre plate reader (Synergy Mx; BioTek, Colmar, France) in accordance with the manufacturer's instructions; the settings were 28°C and constant shaking. Then, every 30 min for 72 h, OD$_{600nm}$ as well as red, yellow and blue fluorescence was measured. The wavelength settings (excitation/emission) were 558 nm/586 nm, 505 nm/530 nm and 435 nm/478 nm for red, yellow and blue fluorescence respectively. Fluorescence was expressed as mean specific fluorescence values (SFU/h, mean value of SFU per hour). Cultures were performed at least in duplicate.

Screening for Lip2 deletion

To screen strains for the Δlip2 phenotype, isolated transformants were grown on 200 μl of YNB for 12 h in a 96-well plate (one colony per well); a drop test was then carried out using a YNB-tributyrin plate supplemented with 10 g l^{-1} of tributyrin. The stock solution of tributyrin (20% tributyltin, 1% Tween) was subjected to 1 min of sonication three times on ice to obtain an emulsion. Colonies and halos could be observed after 2–3 days of culture at 28°C.

Screening for gsy1 deletion

To screen strains for the Δgsy phenotype, isolated transformants were grown on 200 μl of YPD for 12 h in a 96-well plate (one colony per well). The plate was then centrifuged for 5 min at 570 × g. The supernatant was eliminated, and 30 μl of Lugol's solution was added to the plate to stain the pellet. The difference between negatives and positives was visible within 1–2 min. Lugol's solution was prepared by mixing solutions of 2% KI and 1% I$_2$ in a ratio of 1:1.

Screening for xylose utilization

To screen for strains expressing the xylose pathway, isolated transformants were grown in YNB agar plates containing 2% xylose as the sole carbon source. We were able to identify positive colonies after 2–3 days of growth at 28°C.

Acknowledgements

Conflict of interest

The authors declare that they have no competing interests.

References

Beneyton, T., Thomas, S., Griffiths, A.D., Nicaud, J.-M., Drevelle, A., and Rossignol, T. (2017) Droplet-based microfluidic high-throughput screening of heterologous enzymes secreted by the yeast *Yarrowia lipolytica*. *Microb Cell Fact* **16:** 18.

Beopoulos, A., Cescut, J., Haddouche, R., Uribelarrea, J.-L., Molina-Jouve, C., and Nicaud, J.-M. (2009) *Yarrowia lipolytica* as a model for bio-oil production. *Prog Lipid Res* **48:** 375–387.

Beopoulos, A., Verbeke, J., Bordes, F., Guicherd, M., Bressy, M., Marty, A., and Nicaud, J.-M. (2014) Metabolic engineering for ricinoleic acid production in the oleaginous yeast *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **98:** 251–262.

Bhutada, G., Kavscek, M., Ledesma-Amaro, R., Thomas, S., Rechberger, G.N., Nicaud, J.M., and Natter, K. (2017) Sugar versus fat: elimination of glycogen storage improves lipid accumulation in *Yarrowia lipolytica*. *FEMS Yeast Res* **17:** fox020.

Bordes, F., Fudalej, F., Dossat, V., Nicaud, J.M., and Marty, A. (2007) A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. *J Microbiol Methods* **70:** 493–502.

Celinska, E., Ledesma-Amaro, R., Larroude, M., Rossignol, T., Pauthenier, C., and Nicaud, J.M. (2017) Golden Gate Assembly system dedicated to complex pathway manipulation in *Yarrowia lipolytica*. *Microb Biotechnol* **10:** 450–455.

Celińska, E., Borkowska, M., Białas, W., Korpys, P., and Nicaud, J.-M. (2018) Robust signal peptides for protein secretion in *Yarrowia lipolytica*: identification and characterization of novel secretory tags. *Appl Microbiol Biotechnol* **102:** 5221–5233.

Chin, J.X., Chung, B.K.-S., and Lee, D.-Y. (2014) Codon Optimization OnLine (COOL): a web-based multi-objective optimization platform for synthetic gene design. *Bioinformatics* **30:** 2210–2212.

Curran, K.A., Morse, N.J., Markham, K.A., Wagman, A.M., Gupta, A., and Alper, H.S. (2015) Short synthetic terminators for improved heterologous gene expression in yeast. *ACS Synth Biol* **4:** 824–832.

Dulermo, R., Brunel, F., Dulermo, T., Ledesma-Amaro, R., Vion, J., Trassaert, M., *et al.* (2017) Using a vector pool containing variable-strength promoters to optimize protein production in *Yarrowia lipolytica*. *Microb Cell Fact* **16:** 31.

Egermeier, M., Sauer, M., and Marx, H. (2019) Golden Gate based metabolic engineering strategy for wild-type strains of *Yarrowia lipolytica*. *FEMS Microbiol Lett* **366:** fnz022.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009) Golden gate shuffling: a one-pot dna shuffling method based on type IIs restriction enzymes. *PLoS ONE* **4:** e5553.

Groenewald, M., Boekhout, T., Neuvéglise, C., Gaillardin, C., van Dijck, P.W.M., and Wyss, M. (2013) *Yarrowia lipolytica*: safety assessment of an oleaginous yeast with a great industrial potential. *Crit Rev Microbiol* 1–20.

Holkenbrink, C., Dam, M.I., Kildegaard, K.R., Beder, J., Dahlin, J., Belda, D.D., and Borodina, I. (2018) EasyCloneYALI: CRISPR/Cas9-based synthetic toolbox for engineering of the yeast *Yarrowia lipolytica*. *Biotechnol J* **13:** 1700543.

Larroude, M., Celinska, E., Back, A., Thomas, S., Nicaud, J.M., and Ledesma-Amaro, R. (2018) A synthetic biology approach to transform *Yarrowia lipolytica* into a competitive biotechnological producer of beta-carotene. *Biotechnol Bioeng* **115:** 464–472.

Lazar, Z., Rossignol, T., Verbeke, J., Crutz-Le Coq, A.-M., Nicaud, J.-M., and Robak, M. (2013) Optimized invertase expression and secretion cassette for improving Yarrowia lipolytica growth on sucrose for industrial applications. *J Ind Microbiol Biotechnol* **40:** 1273–1283.

Lazar, Z., Liu, N., and Stephanopoulos, G. (2018) Holistic approaches in lipid production by *Yarrowia lipolytica*. *Trends Biotechnol* **36:** 1157–1170.

Le Dall, M.T., Nicaud, J.M., and Gaillardin, C. (1994) Multiple-copy integration in the yeast *Yarrowia lipolytica*. *Curr Genet* **26:** 38–44.

Ledesma-Amaro, R., and Nicaud, J.-M. (2016) Yarrowia lipolytica as a biotechnological chassis to produce usual and unusual fatty acids. *Prog Lipid Res* **61:** 40–50.

Ledesma-Amaro, R., Lazar, Z., Rakicka, M., Guo, Z., Fouchard, F., Coq, A.-M.C.-L., and Nicaud, J.-M. (2016) Metabolic engineering of *Yarrowia lipolytica* to produce chemicals and fuels from xylose. *Metab Eng* **38:** 115–124.

Lee, M.E., DeLoache, W.C., Cervantes, B., and Dueber, J.E. (2015) A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth Biol* **4:** 975–986.

Leplat, C., Nicaud, J.-M., and Rossignol, T. (2015) High-throughput transformation method for *Yarrowia lipolytica* mutant library screening. *FEMS Yeast Res* **15:** fov052-fov052.

Madzak, C. (2018) Engineering *Yarrowia lipolytica* for use in biotechnological applications: a review of major achievements and recent innovations. *Mol Biotechnol* **60:** 621–635.

Niehus, X., Crutz-Le Coq, A.-M., Sandoval, G., Nicaud, J.-M., and Ledesma-Amaro, R. (2018) Engineering *Yarrowia lipolytica* to enhance lipid production from lignocellulosic materials. *Biotechnol Biofuels* **11:** 11.

Park, Y.-K., Korpys, P., Kubiak, M., Celinska, E., Soudier, P., Trébulle, P., et al. (2019) Engineering the architecture of erythritol-inducible promoters for regulated and enhanced gene expression in *Yarrowia lipolytica*. *FEMS Yeast Res* **19:** foy105-foy105.

Pignede, G., Wang, H.-J., Fudalej, F., Seman, M., Gaillardin, C., and Nicaud, J.-M. (2000) Autocloning and amplification of LIP2 in *Yarrowia lipolytica*. *Appl Environ Microbiol* **66:** 3283–3289.

Redden, H., Morse, N., and Alper, H.S. (2015) The synthetic biology toolbox for tuning gene expression in yeast. *FEMS Yeast Res* **15:** 1–10.

Soudier, P., Larroude, M., Celinska, E., Rossignol, T., and Nicaud, J.M. (2019) Selection of heterologous protein-producing strains in *Yarrowia lipolytica*. *Methods Mol Biol* **1923:** 153–168.

Trassaert, M., Vandermies, M., Carly, F., Denies, O., Thomas, S., Fickers, P., and Nicaud, J.-M. (2017) New inducible promoter for gene expression and synthetic biology in *Yarrowia lipolytica*. *Microb Cell Fact* **16:** 141.

Tsakraklides, V., Kamineni, A., Consiglio, A.L., MacEwen, K., Friedlander, J., Blitzblau, H.G., et al. (2018) High-oleate yeast oil without polyunsaturated fatty acids. *Biotechnol Biofuels* **11:** 131.

Verbeke, J., Beopoulos, A., and Nicaud, J.-M. (2013) Efficient homologous recombination with short length flanking fragments in Ku70 deficient *Yarrowia lipolytica* strains. *Biotechnol Lett* **35:** 571–576.

Wong, L., Engel, J., Jin, E., Holdridge, B., and Xu, P. (2017) YaliBricks, a versatile genetic toolkit for streamlined and rapid pathway engineering in *Yarrowia lipolytica*. *Metabol Eng Commun* **5:** 68–77.

Xue, Z., Sharpe, P.L., Hong, S.P., Yadav, N.S., Xie, D., Short, D.R., et al. (2013) Production of omega-3 eicosapentaenoic acid by metabolic engineering of *Yarrowia lipolytica*. *Nat Biotechnol* **31:** 734–740.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Schematic representation of all the bricks available for each position with the corresponding 4-nt overhangs.
Table S1. List of strains and plasmids.
File S1. Fasta sequences of each DNA module.
Data S1. Protocol for assembling a multigene pathway using the Golden Gate toolkit for *Yarrowia lipolytica*.

To highlight the importance of the developed system in the construction of complex strains, the heterologous biosynthetic pathway for carotenoids biosynthesis was assembled and expressed in *Y. lipolytica*. The pathway, composed of genes *CarB* (phytoene dehydrogenase) and *CarRP* (phytoene synthase) from *Mucor circinelloides*, was assembled with a copy of the endogenous gene *GGS1* (geranylgeranyl diphosphate synthase; YALI0D17050g) in one plasmid using the GG 3TU-approach. Plugged on the mevalonate pathway, through farnesyl pyrophosphate, the heterologous pathway allows the production of β-carotene. This work was published in two consecutives and complementary articles (Celińska et al., 2017; Larroude et al., 2017). The first one described the GG system for the first time in *Y. lipolytica*, outlining the development of the thirteen 4-nt overhangs scaffold strategy that characterize the GG system developed for this yeast. During this work the set of building blocks available for each position was still reduced, however the efficiency and capacity of the system was shown by producing carotenoids in two strains of *Y. lipolytica*. When expressing this three gene pathway, 67-90% of the obtained clones exhibited the desired phenotype, improving the 20% efficiency obtained when the same pathway was assembled and expressed by other one-step method (Gao et al., 2014). The construction of the first elements of the GG library was part of the present thesis work. This published article can be found on the annexes of this manuscript (6.2).

The second work is the exploitation of the system to construct a cell factory and produce a molecule of interest. The GG method was used to construct complex strains in a much more efficient, fast and hence cheaper way. Using this technique, the heterologous pathway to synthesise β-carotene was easily expressed in a wild-type strain and in a strain optimized for the synthesis of lipids. The rapid construction and integration of the entire pathway allowed its easy expression on different strains. This permitted to quickly evidence that the lipid producer strain was a better chassis strain for the production of β-carotene, producing 0.035 g/L of β-carotene compared to 0.018 g/L in the wild-type strain. The rapid integration of the pathway also permitted the simple expression of additional copies of the pathway into the strain, increasing the production to 0.280 g/L when one extra copy was added and to 0.450 g/L when the pathway was expressed three times in the cell. The best β-carotene producer strain obtained was tested in different media, achieving the highest production, 1.5 g/L, with YPD and 60g/L of glucose. Finally, the production was boosted by scaling up the culture from

250 mL flask to a 5 L bioreactor with optimized and controlled conditions, reaching 6.5 g/L of β-carotene after 122 h of culture, the highest production reported so far.

During this work, a combinatorial approach, based on the GG strategy, was developed and used to screen the optimum promoter-gene pairs for each TU expressed for the carotenoid pathway (Figure 15). To do so, several building blocks assigned to the same GG position are added to the assembly mix allowing a shuffling of these parts, in the case here, the promoters. After the digestion-ligation cycles a pool of cassettes is obtained and used for transformation. The best combination is identified *a posteriori*. In our case, this pathway optimization increased more than 40% the production of β-carotene. It has been shown that every possible combination is unbiased generated during the combinatorial assemblies (Awan et al., 2017). Thus, this shuffling parts capacity expands the uses of GG strategy.

P1	P2	P3	colour
TEF	TEF	TEF	+++
TEF	GAP	PGM	++
TEF	GAP	GAP	++
PGM	TEF	GAP	++
PGM	TEF	PGM	++
PGM	TEF	TEF	++
PGM	GAP	GAP	+
GAP	TEF	GAP	+
GAP	TEF	TEF	+
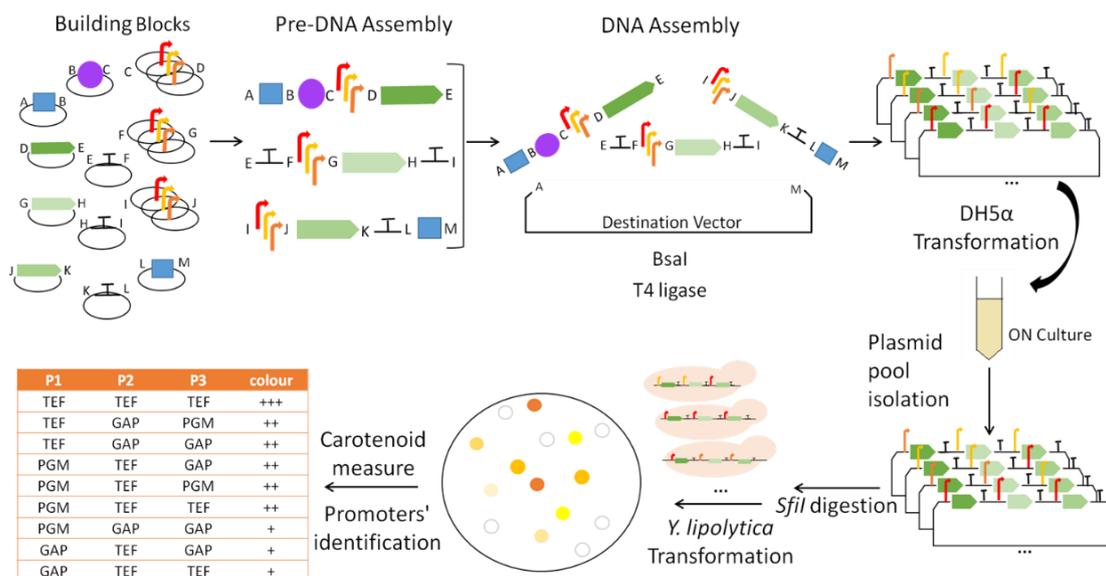
Figure 15. Combinatorial approach using Golden Gate. Example of a promoter shuffling strategy.

Briefly, GG cloning system is used to generate a pool of randomized expression cassettes. The pool of expression cassettes is used to transform *Y. lipolytica* and the generated clones are afterwards screened. The selected clones are then analysed to identify the construction that lead to the desired phenotype. In the presented example, the pool contained constructions with different promoters and the strains were screened by their capacity to produce β-carotene. Figure from Larroude et al., 2017

In addition, this work highlights the importance of selecting the right microorganism and strain in a metabolic engineering project. The selection of an oleaginous organism permitted higher production levels than with other yeasts such as *S. cerevisiae*, while the selection of a lipid overproducer strain further increases the β-carotene content.

The complete results are presented in the article published in Biotechnology and Bioengineering presented hereafter (Larroude et al., 2017)

ARTICLE

WILEY BIOTECHNOLOGY BIOENGINEERING

A synthetic biology approach to transform *Yarrowia lipolytica* into a competitive biotechnological producer of β-carotene

Macarena Larroude[1] | Ewelina Celinska[2] | Alexandre Back[1] | Stephan Thomas[1] | Jean-Marc Nicaud[1] (iD) | Rodrigo Ledesma-Amaro[1,3] (iD)

[1] BIMLip, Biologie Intégrative du Métabolisme Lipidique Team, Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

[2] Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, Poznan, Poland

[3] Department of Bioengineering, Imperial College London, London, UK

Correspondence
Rodrigo Ledesma-Amaro, Department of Bioengineering, Imperial College London, London, UK.
Email: rodrigoledesma@usal.es

Abstract

The increasing market demands of β-carotene as colorant, antioxidant and vitamin precursor, requires novel biotechnological production platforms. *Yarrowia lipolytica*, is an industrial organism unable to naturally synthesize carotenoids but with the ability to produce high amounts of the precursor Acetyl-CoA. We first found that a lipid overproducer strain was capable of producing more β-carotene than a wild type after expressing the heterologous pathway. Thereafter, we developed a combinatorial synthetic biology approach base on Golden Gate DNA assembly to screen the optimum promoter-gene pairs for each transcriptional unit expressed. The best strain reached a production titer of 1.5 g/L and a maximum yield of 0.048 g/g of glucose in flask. β-carotene production was further increased in controlled conditions using a fed-batch fermentation. A total production of β-carotene of 6.5 g/L and 90 mg/g DCW with a concomitant production of 42.6 g/L of lipids was achieved. Such high titers suggest that engineered *Y. lipolytica* is a competitive producer organism of β-carotene.

KEYWORDS
β-carotene, golden gate, metabolic engineering, promoter shuffling, synthetic biology, *Yarrowia lipolytica*

1 | INTRODUCTION

β-carotene is an orange pigment, precursor of vitamin A. This compound is a biochemical synthesized terpenoid that belongs to the group of carotenoids, together with lycopene, canthaxanthin, astaxanthin among others. Carotenoids have antioxidants properties which make them very relevant industrial compounds, with an expected market of $1.4 billion in 2018 (Lin, Jain, & Yan, 2014; Ye & Bhatia, 2012). β-carotene is produced either chemically or biotechnologically using natural producer microorganisms such as *Blakeslea trispora* (Nanou & Roukas, 2016), *Xanthophyllomyces dendrorhous* (Contreras et al., 2015) or *Dunaliella salina* (Wichuk, Brynjolfsson, & Fu, 2014). However, the heterologous production through metabolic engineering is considering a promising way to optimize β-carotene synthesis and face the increasing market

demands. Therefore, several strategies to produce this compound have been carried out, mainly in model organisms such as *Saccharomyces cerevisiae* and *Escherichia coli*. Such strategies include the expression of heterologous genes, the elimination/downregulation of competing pathways, the overexpression of endogenous genes, adaptive evolution, fine control of the metabolic fluxes, etc. (Lin et al., 2014; Ye & Bhatia, 2012).

To our knowledge, the best production of β-carotene achieved in engineered baker yeast reached a production of 0.45 g/L and 8.12 mg/g DCW (Xie, Ye, Lv, Xu, & Yu, 2015). An adaptive evolution approach generated the strain with the highest relative production; 18 mg/g DCW (Reyes, Gomez, & Kao, 2014).

Higher values have been obtained by engineering *E. coli*; 3.2 g/L from glycerol (Yang & Guo, 2014) or 2.47 g/L from glucose (Nam, Choi, Lee, Kim, & Oh, 2013). One of the main drawback of using this bacteria

for a commercial production of β-carotene is the safety issues associated to the lack of the GRAS status.

Yarrowia lipolytica is an oleaginous yeast widely investigated and modified for the production of biotechnologically relevant compounds (Ledesma-Amaro & Nicaud, 2015; Liu, Ji, & Huang, 2015; Madzak, 2015; Zhu & Jackson, 2015). It presents several advantages as industrial host, including a vast repertoire of molecular tools and the ability to grow naturally in low cost substrates such as glycerol or molasses (Ledesma-Amaro & Nicaud, 2016), or once engineered in xylose, raw starch, cellobiose, cellulose, or inulin (Guo et al., 2015; Ledesma-Amaro, Dulermo, & Nicaud, 2015; Ledesma-Amaro, Lazar et al., 2016; Wei et al., 2014; Zhao, Cui, Liu, Chi, & Madzak, 2010). Moreover, its metabolism, specifically the lipid metabolism, has been widely studied and characterized (Dulermo, Gamboa-Melendez, Ledesma-Amaro, Thevenieau, & Nicaud, 2015; Dulermo, Gamboa-Melendez, Ledesma-Amaro, Thevenieau, & Nicaud, 2016; Kerkhoven et al., 2017; Kerkhoven, Pomraning, Baker, & Nielsen, 2016; Qiao, Wasylenko, Zhou, Xu, & Stephanopoulos, 2017; Wasylenko, Ahn, & Stephanopoulos, 2015). Interestingly, lipid biosynthetic pathway and carotenoid pathway share a common precursor, Acetyl-CoA, which is highly available in *Y. lipolytica*. In addition, the genome of this yeast encodes all the required genes to produce geranylgeranyl diphosphate (GGPP), only two compounds away from β-carotene (Figure 1). Because of this, *Y. lipolytica* has been recently proposed as a potential producer of carotenoids. Both DuPont and Microbia have patented modified strains of *Y. lipolytica* engineered to produce carotenoids (Bailey, Madden, & Trueheart, 2012; Sharpe, Ye, & Zhu, 2014). Interestingly, the authors proved that modified *Y. lipolytica* was able to produce lycopene (2 mg/gDCW), β-carotene (5.7 mg/gDCW), canthaxanthin and astaxanthin (Sharpe et al., 2014). Another metabolic engineering approach produced 16 mg/gDCW of lycopene, the direct precursor of β-carotene

(Matthaus, Ketelhot, Gatter, & Barth, 2014). Importantly, Grenfell-Lee, Zeller, Cardoso, and Pucaj (2014) have demonstrated that the safety profile of the β-carotene produced in *Y. lipolytica* is the same as other commercial products, which would facilitate its commercialization. Therefore, this oleaginous yeast represents a promising biotechnological chassis for the production of β-carotene.

In addition, during the preparation of this manuscript, Gao et al. (2017) engineered *Y. lipolytica* to produce up to 4 g/L of β-carotene in fed-batch fermentation; the highest production titer so far described in a heterologous microorganism. The authors engineered one strain after 12 steps where 11 genes were modified and they found that the integration of multiple copies of some of the genes was essential to increase β-carotene production.

In this work, we first found that a strain that overproduce lipids is more convenient for the production of β-carotene than a wild-type background. We further engineered the lipid overproducer *Y. lipolytica* strain in order to maximize β-carotene production. For this aim, we used a synthetic biology approach in order to screen the best combination of promoters for each of the studied genes. Finally, we integrated the metabolic engineering of the strain with fermentation condition optimization in order to boost β-carotene production, reaching the best production titer and yield described so far.

2 | MATERIAL AND METHODS

2.1 | Strains and media

Escherichia coli strain DH5α was used for cloning and plasmid propagation. Cells were grown at 37°C with constant shaking on 5 ml LB medium (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl), and ampicillin (100 μg/ml) or kanamycin (50 μg/ml) were added for plasmid selection.

The *Y. lipolytica* strains used in this study are derived from Po1d (wt), derived from the wild-type *Y. lipolytica* W29 (ATCC20460) strain. All the strains used in this study are listed in Supplementary Figure S1. Media and growth conditions for *Y. lipolytica* have been described elsewhere (Dulermo et al., 2015). Rich media YPD and YPD60 contained 1 or 6% glucose (Sigma–Aldrich, Saint-Quentin-Fallavier, France) respectively, 1% peptone (BD Bioscience, Le Pont de Claix, France) and 1% yeast extract (BD Bioscience). Minimal media YNB20, YNB30, and YNB60, contained 2, 3, or 6% glucose respectively (wt/vol; Sigma), 0.17% (wt/vol) Yeast Nitrogen Base (YNBww; Difco), 0.5% (wt/vol) NH_4Cl and 50 mM phosphate buffer (pH6.8). YNBgly60 medium was prepared in the same way as the YNB medium except that 6% glycerol was added as sole carbon source. When necessary, the YNB medium was supplemented with uracil (0.1 g/L) and/or leucine (0.1 g/L) or hygromycin (0.2 g/L). Solid media for *E. coli* and *Y. lipolytica* was prepared by adding 15 g/L agar (Thermo Fisher Scientific, Courtaboeuf, France) to liquid media.

2.2 | Strains construction

All restriction enzymes were purchased from New England Biolabs (NEB). PCR amplifications were performed using Q5 high-fidelity DNA
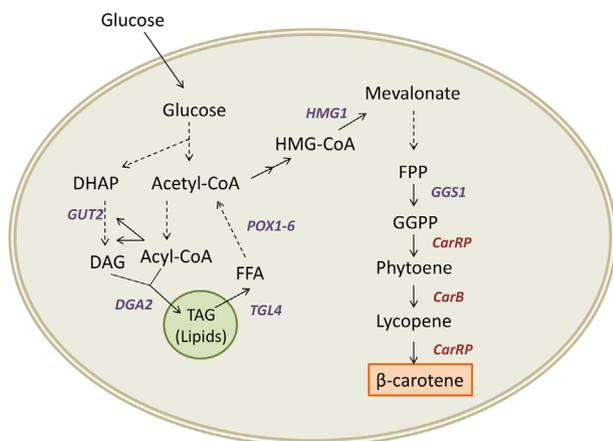
FIGURE 1 Scheme of the metabolic pathways leading to the production of lipids and β-carotene in *Y. lipolytica*. Metabolites are shown in black, native enzymes in purple and heterologous enzymes in red. The continuous arrows indicate a metabolic step while the dashed arrows indicate multiple metabolic steps. DHAP, dihydroxyacetone phosphate; DAG, diacylglycerol; TAG, triacylglycerol; FFA, free fatty acids; FPP, farnesyl pyrophosphate; GGPP, geranylgeranyl pyrophosphate

polymerase (NEB) or GoTaq DNA polymerase (Promega, Charbon-nières-les-Bains, France). When needed PCR fragment were purified using the QIAquick Gel Extraction Kit (Qiagen, Courtaboeuf, France). Plasmids from *E. coli* were extracted using the QIAprep Spin Miniprep Kit (Qiagen). All the reactions were performed according to the manufacturer instructions. Transformation of chemically competent *E. coli* cells was performed by thermic shock protocol. Transformation of *Y. lipolytica* was performed using the lithium-acetate method adapted from (Barth & Gaillardin, 1996). Transformants were selected on YNBLeu, YNBura, or YNBHygro media, depending on their genotype.

2.3 | DNA constructions

Primer's sequences used in this study can be found in Supplementary Table S1. The carotenoid's (*GGS1-CarB-CarRP*) expression cassette (Supplementary Figure S2) named car-cassette and the t-HMG's expression cassette were kindly provided by Microbia (Bailey et al., 2012) onto plasmids containing the *LEU2* gene and the nourseothricin resistance gene respectively as selection markers. All sequencing processes were done by Eurofins genomics. A complete plasmid list can be found in Supplementary Table S2.

The construction of carTEF-cassette was done using the recently developed Golden Gate (GG) toolbox for *Y. lipolytica* (Celinska et al., 2017) based on (Engler & Marillonnet, 2014). Briefly, primers carrying pre-designed 4-nt overhangs and externally located BsaI recognition sites were designed and used to amplify the building blocks which were then cloned in donor vectors (Zero Blunt® TOPO® PCR Cloning Kit, Thermo Fisher Scientific). These building blocks and the destination vector carrying a gene encoding for a chromophore, to facilitate selection, which is also flanked with BsaI-site and pre-designed overhangs, are then mixed equimolarly (50 pmoles of ends) in one-pot reaction together with 5 U of BsaI (NEB), 200 U of T4 ligase (NEB), 2 μl of T4 DNA ligase buffer (NEB) and ddH$_2$O up to 20 μl. The following thermal profile was applied: [37°C for 5 min, 16°C for 2 min]×60, 55°C for 5 min, 80°C for 5 min, 15°C ∞. Subsequently, the reaction mixture was used for *E. coli* transformation. White colonies were screened for identification of complete Golden Gate Assembly (GGA) through plasmid isolation, restriction digestion and PCR. Complete GGA was subsequently linearized by SfiI restriction enzyme and 10 μl were used for transformation of *Y. lipolytica*.

All the sequences to be used as building blocks of the envisioned GG Assembly were extracted from *Y. lipolytica* W29 genome sequence or from previously constructed vectors of our own collection. After cloned in TOPO vector, the building blocks were screened by restriction digestion and verified by sequencing.

A promoter shuffling was carried out to explore the combina-tion of multiple promoters (Engler & Marillonnet, 2013). The assembly of plasmids with three randomized promoters (PGMp-low expression, GAPDHp-medium expression, TEFp-high expression) was split into two parts: pre-assembly construct and then multigene construct assembly. For the pre-assembly, three separate reactions with four Golden Gate parts each (InsertionSiteUp-Marker-Promoter1-Gene1;Terminator1-Promoter2-Gene2-Terminator2;

Promoter3-Gene3-Terminator3-InsertionSIteDown) was done. All parts were used in equimolar quantities (50 pmoles), and a mix of all three promoters was made with a final concentration of 50 pmoles. Each reaction has 5 U of BsaI, 200 U of T4 ligase, 1 μl of T4 DNA ligase buffer (NEB) and ddH$_2$O up to 10 μl and a short thermal profile was applied: [37°C for 3 min, 16°C for 2 min]×30, 55°C for 5 min, 80°C for 5 min, 15°C ∞. Subsequently, for the multigene construct assembly the three previous reactions were mixed together in the same tube and the destination vector (50 pmoles) was added together with 20 U of BsaI, 400 U of T4 ligase, 4 μl of T4 DNA ligase buffer (NEB) and ddH$_2$O up to 40 μl. The following thermal profile was applied: [37°C for 5 min, 16°C for 5 min]×50, 55°C for 5 min, 80°C for 5 min, 15°C ∞. This Golden Gate reaction was transformed into *E. coli*, and all transformant colonies were mixed together into a single overnight culture. A plasmid library was prepared from this overnight culture, digested with SfiI restriction enzyme and 10 μl were used to transform *Y. lipolytica* (Figure 4).

Identification of promoters at each site was done by PCR from gDNA of the selected yellow-orange colonies obtained. For promoter in position 1 the primer pair used was Ura3Marker-intern-Fw/GGSI_intern_Rv, for position 2 GGSI_intern_Fw/CarB_intern_Rv, and for position 3 CarB_intern_Fw/CarRP_intern_RV.

2.4 | β-carotene measurement

Intracellular β-carotene content was extracted and quantified by photometric measurement adapted from previous reports (Matthaus et al., 2014). Briefly, 20 ml of medium was inoculated with 0.05 DO pre-culture yeast strain in a 250 ml flask. Cells were cultured during 4 days at 28°C and shaking. Afterwards 200 μl of the culture were harvested in a FastPrep FP120 (Thermo Fisher Scientific) and 500 μl glass beads (0.75 to 1 mm; Roth) and 1.2 ml extraction solvent (50:50 v/v; hexane-ethyl acetate; 0.01% butyl hydroxyl toluene) were added. The mixture was vortexed three times for 1min30s at maximum speed, alternating with ice incubation. The extract was collected after 5 min centrifugation, and the extraction procedure was repeated until the pellet and the supernatant were colorless. The extract was then diluted with extraction solvent and measured photometrically at 448 nm. The concentrations were calculated through a standard curve using β-carotene from Sigma–Aldrich (St. Louis, MO) as standard. The OD was correlated with the dry cell weight (DCW) measurement for each corresponding culture. The washed and lyophilized cells coming from a known volume served to measure the DCW. Every sample was cultured in duplicate.

2.5 | Microscopy images

Images were acquired using a Zeiss Axio Imager M2 microscope (Zeiss, Marly-le-Roi, France) with a 100 × objective and Zeiss filters 45 and 46 for fluorescent microscopy. Axiovision 4.8 software (Zeiss) was used for image acquisition. Lipid bodies visualization was performed by adding BodiPy® Lipid Probe (2.5 mg/ml in ethanol; Thermo Fisher Scientific) to the samples and after incubation at room temperature for 10 min.

Microscopic color images were acquired using a Leica DM1000 microscope (Leica) with a 100× objective a moticam 2500 camera. Motic imaging 2.0 software was used for image acquisition.

2.6 | Lipid content quantification

Lipids were extracted from 15 to 25 mg aliquots of lyophilized cells and converted into their equivalent methyl esters as previously described (Ledesma-Amaro, Dulermo, Niehus, & Nicaud, 2016). The products were then used in the gas chromatography (GC) analysis, performed using a Varian 3900 gas chromatograph equipped with a flame ionization detector and a Varian FactorFour vf–23 ms column, where the bleed specification at 260°C was 3 pA (30 m, 0.25 mm, 0.25 μm). FA were identified by comparison to commercial fatty acid methyl ester standards (FAME32, Supelco) and quantified by the internal standard method with the addition of 50 μg of commercial C12:0 (Sigma).

The washed and lyophilized cells coming from a known volume served to measure the DCW.

2.7 | Biomass, sugar, and acid quantification

Dry cell weight (DCW) was calculated by weighting the lyophilized cells. The harvested cells were washed twice and centrifuged in order to separate all the mass remaining in the culture media prior to lyophilization.

Sugar and citric acid were quantified by HPLC (UltiMate 3000, Thermo Fisher Scientific) using an Aminex HPX 87 H column coupled to UV (210 nm) and RI detectors. The column was eluted with 0.01 N H_2SO_4 at room temperature and a flow rate of 0.6 ml min^{-1}. Identification and quantification were achieved via comparisons to standards. Before being subject to HPLC analysis, samples were filtered on 0.45-μm pore-size membranes.

2.8 | Bioreactor procedures

Fed-batch cultivations were performed in a 5 L bioreactor (Sartorius Stedim Biotech, Göttingen, Germany) with an initial working volume of 2 L. The initial medium (Y10P20D) contained 10 g/L of Yeast extract (Becton Dickinson, Le Pont-de-Claix, France), 20 g/L of peptone (Becton Dickinson) and 5 g/L of glucose (Sigma–Aldrich). The concentrated medium (Y20P40D) contained 20 g/L of Yeast extract, 40 g/L of peptone and 5 g/L of glucose. The temperature was held constant at 28°C, the aeration at 2 VVM, the agitation at 500–900 rpm, the pH at 5.5 automatically by injection of 100 g/L H3PO4 or 200 g/L KOH, and the dissolved oxygen was set up at 20%. Foam was prevented by the addition of antifoam 204 (Sigma–Aldrich). The medium was inoculated with 100 ml from a 24 hr preculture performed in a shake flask containing 10 g/L of Yeast extract, 20 g/L of peptone and 20 g/L of glucose. The fed-batch process was initiated after 6 hr of cultivation at a rate of 6 g/L (0.2 molcarbon/hr) of glucose from a 500 g/L concentrated stock solution.

3 | RESULTS AND DISCUSSION

3.1 | Lipid overproducer strain synthesizes higher amount of β-carotene

Previous reports have proven that the overexpression of three genes, geranylgeranyl diphosphate synthase (*GGS1* from *Y. lipolytica*), pytoene synthase/lycopene cyclase and phytoene dehydrogenase (*carPR* and *carB* from *Mucor circinelloides*) promote β-carotene production in *Y. lipolytica* (Celinska et al., 2017; Gao et al., 2014). We here used an expression cassette (car-cassette) where the expression of *GGS1* is controlled by the promoter PGMp, of *CarB* by GAPDHp and of *CarRP* by TEF1p. As expected, the sole expression of this cassette in the parental strain (named wt-C) allows it to produce substantial amounts of β-carotene (3.4 mg/gDCW and 18.4 mg/L) (Figure 2). This production level was higher than the 2.2 mg/gDCW previously obtained for a similar approach (Gao et al., 2014). The differences could be caused by the use of different promoters controlling each of the three genes or due to the use of different parental strains.

It is well known that the lipophilic nature of carotenoids promotes their storage in the lipid bodies of the cells. In Supplementary Figure S3, the co-localization of the pigment and the neutral lipids can be seen. Here, we hypostatized that lipid overproducer strains could
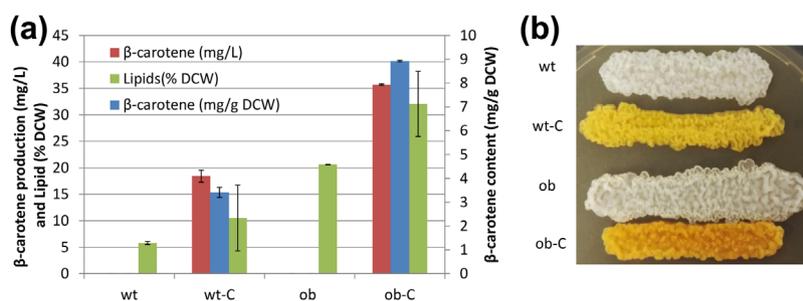
FIGURE 2 Lipids and β-carotene production in wt and ob backgrounds. (a) β-carotene production (mg/L), content (mg/g DCW) and lipid content (% DCW) in the strains wt, wt-C, ob, and ob-C after growing in YPD for 4 days. The average value and the SD from two independent experiments are shown. (b) Strains wt, wt-C, ob, and ob-C growing on YPD-agar plate where the color provoked by the β-carotene can be seen
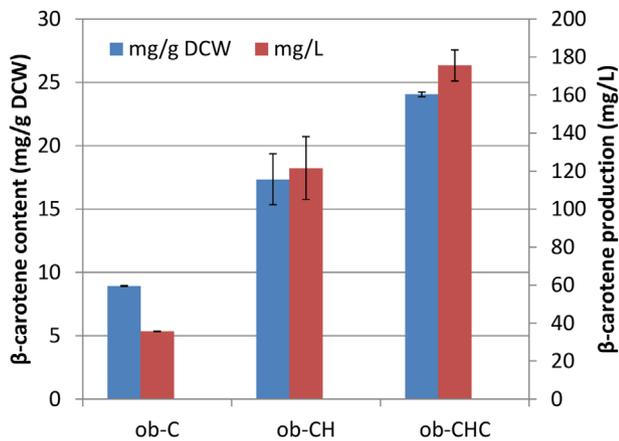
FIGURE 3 β-carotene production in ob-derived background. (a) β-carotene production (mg/L) and content (mg/g DCW) in the strains ob-C, ob-CH, and ob-CHC after growing in YPD for 4 days. The average value and the SD from two independent experiments are shown

boost both β-carotene production, due to a higher availability of the precursor Acetyl-CoA, and storage, due to an increase of the lipophilic structures inside the cells. In order to verify this, we transformed a lipid overproducer strain (JMY3501, hereafter called "obese" or "ob") with the car-cassette. As expected, the generated strain (ob-C) accumulated higher amount of lipids, 3.6 times more than the strain wt-car, and interestingly it also boosted β-carotene synthesis, producing 8.9 mg/gDCW and 35.7 mg/L, 2.61 and 1.93 times more than wt-car (Figure 2). The higher amounts of lipids could be important for the solubilization of β-carotene, which would reduce the formation of crystals that could impair the cellular homeostasis. These results indicate the importance of selecting the proper parental strain in metabolic engineering approaches. Unexpectedly, the expression of the β-carotene pathway favored the total production of lipids, which increased 82% and 56% in

the wild type and obese background respectively. Importantly, the co-production of the two biotechnologically relevant compounds, β-carotene and lipids, could facilitate the industrial viability of the process.

3.2 | Further metabolic engineering increases carotene content and reveals pathway bottlenecks

The overexpersion of hydroxymethylglutaryl-CoA reductase (HMG1) is known to channelized the flux toward carotenoids in engineered microorganisms (Matthaus et al., 2014). Here, we overexpressed a truncated version of this gene (YALI0E04807, [Bailey et al., 2012]) under the control of the constitutive TEF promoter in the strain ob-C. The generated strain (ob-CH) increased β-carotene content up to 17.4 mg/gDCW and 121.6 mg/L, 2.0 and 3.4 times more than the parental strain ob-C (Figure 3). Again, this results are higher than the ones recently obtained from a strain overexpressing the same genes but using different parental strain and set of promoters (4.36 mg/gDCW and 64.6 mg/L) (Gao et al., 2017).

In order to analyze if the production of β-carotene is limited by the expression levels of the overexpressed genes we introduced in the genome of ob-CH a second copy of the car-cassette, generating the ob-CHC. Interestingly, we found a further increase in the desired product, which reached 24.0 mg/gDCW and 175.6 mg/L, 1.4 and 1.4 times more than the parental strain ob-CH. This result indicates that the expression of some of the gene encoded in the car-cassette is limiting the production of β-carotene. Similar conclusion was reached by Gao et al. (2017) in a parallel study where the authors improved the β-carotene content by increasing the copy number of the genes expressed. Here, in order to overcome this limitation we considered a strategy that combines the increase of the copy number and the optimization of the promoter strength for each gene.
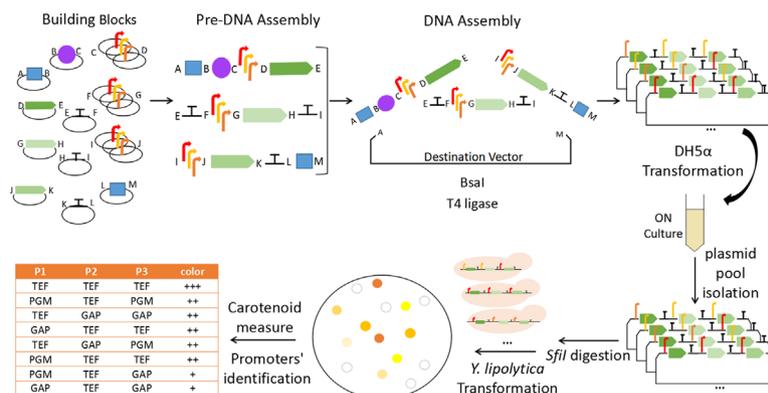
FIGURE 4 Scheme of the promoter shuffling strategy to optimize β-carotene production in *Y. lipolytica*. Briefly, Golden Gate cloning system was used to generate a pool of randomized expression cassettes bearing different set of three different promoters. The pool of expression cassettes was used to transform *Y. lipolytica* and the generated clones were screened by color intensity. The selected clones were further analyzed, each promoter set was determined and the β-carotene content was measured (see section 2 for a detailed explanation). Letters A–M represent the designed overhangs enabling the ordered assembly of DNA parts after BsaI digestion. Blue squares represent genomic integration targeting sequences, violet circle represents the selection marker gene, red–orange–yellow arrows represent promoters, green arrows represent genes, and T represent terminators
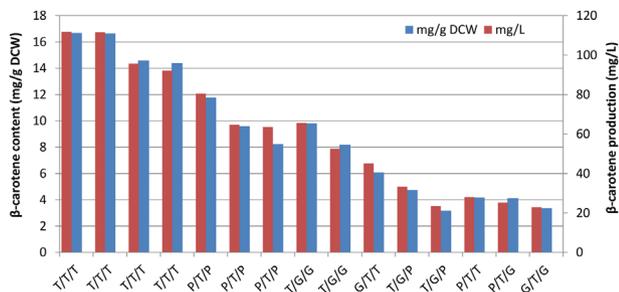
FIGURE 5 β-carotene production in the strains selected from the promoter shuffling approach. β-carotene production (mg/L) and content (mg/g DCW) in the strains derived from the promoter shuffling experiment after growing in YPD for 4 days. Each strain is identified by the set of promoters: T, TEFp; P, PGMp; G, GAPDHp in the car-cassette order 1st GGS1, 2nd carB, and 3rd carRP. Some of the selected strains presented the same combination of promoters

3.3 | Promoter shuffling using Golden Gate indentifies the best promoter set for the production of β-carotene

We have previously seen (Dulermo et al., 2017) that the increase in the promoter strength, even in strong promoters, can enhance the transcription level more than six times, while an extra copy of a gene under the same promoter can typically only duplicate the expression level. We therefore decided to identify the best combination of promoters and genes in order to maximize β-carotene production. The construction of the car-cassette with different promoters via traditional cloning would have been a very long and inefficient process. Thus, we took advantage of the recently developed Golden Gate toolbox for *Y. lipolytica* (Celinska et al., 2017) to perform a promoter shuffling strategy. The strategy, summarized in Figure 4, consists in a digestion-ligation reaction guided by BsaI defined sites where the three promoters can be introduced in each of the three promoter positions in the car-cassette. It has been recently proven by sequencing the Golden Gate reaction products that every possible combination is unbiased generated during

combinatorial assemblies (Awan et al., 2017). Then, the pool of cassettes were simultaneously amplified, linearized and used to transform the wild type parental strain (wt), where the difference in color intensity could be easily screened (Supplementary Figure S4). After the transformation, 387 colonies with different color were obtained. Here, we selected the 15 more orange–yellow strains and we checked the set of promoter controlling each gene (Figure 5). Interestingly, the combination of promoters originally present in the car-cassette was not found among the best producer strains and, accordingly, the analyzed strains generated by the shuffling produced higher β-carotene content than the wt-car previously analyzed (Figure 2). The major conclusion of these results is that the cassette with the three genes controlled by TEF1p is the optimum producer. Analyzing the overall combination of promoters found as convenient for overproducing β-carotene, it seems that the presence of a strong TEF promoter is favored in the second position of the car-cassette (11 out of 15). The best of these strains were able to produce 6.3 times more β-carotene than the wt-car, 16.7 mg/gDCW and 111.8 mg/L.

Based on these results we constructed using Golden Gate a new car-cassette where the three genes, *GGS1*, *CarB* and *CarRP* are under the control of TEF1 promoter. This new cassette was called car^{TEF}-cassette and it was used to further increase the β-carotene content of ob-CH.

3.4 | Construction of a β-caratone overproducer strain

We have previously seen that an extra copy of the car-cassette in the ob-CH increased β-carotene production and, in addition, we have improved the cassette (car^{TEF}-cassette) by an optimized set of promoters. We therefore attempted to further increase total β-carotene content by the expression of car^{TEF}-cassette in the strain ob-CH. The generated strain, ob-CHC^{TEF}, as hypothesized, produced higher β-caratone; 54.4 mg/gDCW and 293.3 mg/L (Figure 6).

Finally, an extra copy of the car^{TEF}-cassette was introduced to construct the strain ob-CHC^{TEF}C^{TEF}. This strain was able to
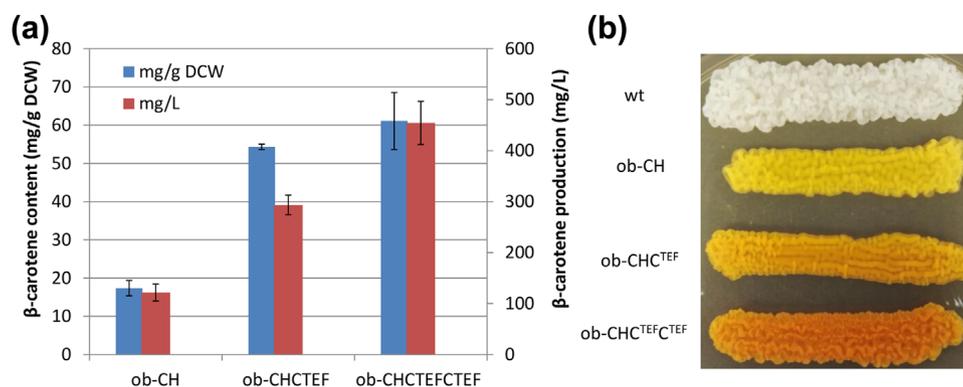
FIGURE 6 β-carotene production in engineered strains of *Y. lipolytica*. (a) β-carotene production (mg/L) and content (mg/g DCW) in the strains ob-CH, ob-CHCTEF, and ob-CHCTEFCTEF after growing in YPD for 4 days. The average value and the SD from two independent experiments are shown. (b) Strains wt, ob-CH, ob-CHCTEF, and ob-CHCTEFCTEF growing on YPD-agar plate where the color provoked by the β-carotene can be seen
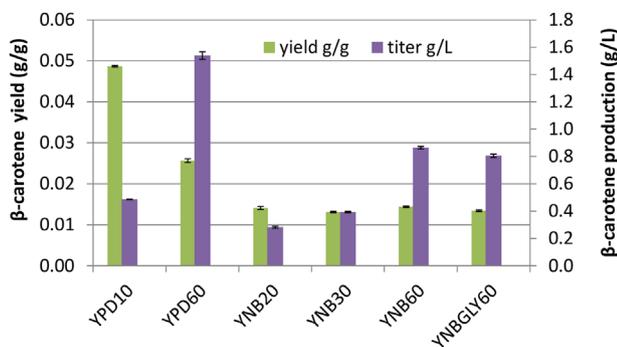
FIGURE 7 Yield and titer of β-carotene production in different culture media. (a) β-carotene yield (g β-carotene/g substrate) and titer (g β-carotene/L) in the strain ob-CHCTEFCTEF after growing for 4 days in rich media (YPD10, 10g/L glucose or YPD60, 60g/L glucose) or synthetic media (YNB20, 20g/L glucose, YNB30, 30g/L glucose, YNB 60, 60 g/L glucose or YNBGLY60, 60 g/L glycerol). The average value and the SD from two independent experiments are shown

produce 61.1 mg/gDCW and 454.36 mg/L of β-carotene (Figure 6), which represent the best production of this compound so far described in flask culture (260 mg/L in _E. coli_ (Yang & Guo, 2014) and 353.6 mg/L in _Y. lipolytica_ (Gao et al., 2017) and the best yield reported so far (0.048 g/g followed by 0.027 g/g achieved in _E. coli_ (Yang & Guo, 2014) and 0.018 g/g in _Y. lipolytica_ (Gao et al., 2017)). We also tested the stability of this strain, since some reports in _S. cerevisae_ showed the appearance of spontaneous white colonies (Beekwilder et al., 2014), however, in _Y. lipolytica_ all the cells plated after 6 days of culture produced β-carotene (Supplementary Figure S5), which could be expected from the low homologous recombination rate in this yeast. We therefore selected this strain in order to optimize culture conditions to further increase β-carotene production.

3.5 | Media optimization shows a trade-off between production titer and yield

In order to study the effect of the media composition in the production of β-carotene, we tested two different kind of media, rich media (YPD) and synthetic media (YNB). We also tested different concentrations of carbon source keeping constant the nitrogen amount, referred by a number that indicates the concentration in g/L (10, 20, 30, and 60). It is well known that a higher C/N ratio promotes lipid production as well as other carbon based molecules such as carotenoids (Braunwald et al., 2013). We tested glucose as a standard carbon source and glycerol (GLY) as a cheaper carbon sources. The selected media were YPD10, YPD60, YNB20, YNB30, YNB60, and YNBGLY60. Large variations in the β-carotene production were found depending on the culture media (Figure 7). According to the results we can suggest that there is no much influence on the carbon source since glucose and glycerol showed similar titer and yields. In addition, a clear correlation between the increase in the initial glucose content and the production titer was found in both rich and synthetic media. Interestingly, the production yields for all the YNB based-media tested was similar independently of the amount or kind of carbon source. However, this was not the case for rich media where a trade-off between production titer and yield was found. The best β-carotene titer so far, 1.5 g/L, was found in YPD60 while the best yield, 0.048 g/g was produced in YPD10. In any case, YPD showed higher titers and yields than YNB and therefore we selected rich media for further optimization of the culture conditions in a controlled fermentation in bioreactor.

3.6 | Bioreactor controlled conditions boosts β-carotene production by the engineered strain ob-CHC^{TEF}C^{TEF}

In order to improve the production of β-carotene in the strain ob-CHC^{TEF}C^{TEF} we decided to optimize the culture conditions using a 5 L
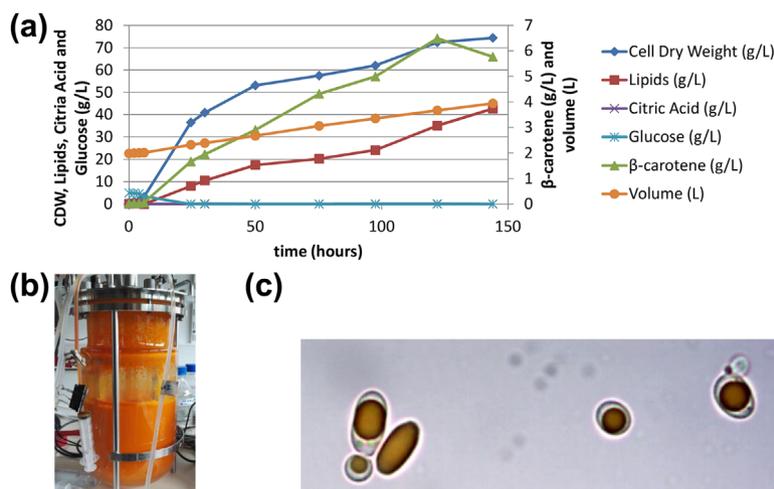
FIGURE 8 Production of β-carotene by engineered _Y. lipolytica_ in bioreactor in Y20P40D media. (a) Kinetics of the 5-L bioreactor fermentation along 143.5 hr showing Cell Dry Weight (g/L), lipids (g/L), citric acid (g/L), glucose (g/L), β-carotene (g/L), and bioreactor broth volume (L). (b) Picture of the bioreactor after 120 hr in Y10P20D. (c) microscopic image of the cells grown in Y10P20D where the β-carotene can be seen as the orange color staining the lipid bodies of the cells

bioreactor. We here performed a fed-batch fermentation using rich media (Y10P20D) where glucose was added after 6 h of cultivation at a rate of 6 g/h. The fermentation was stopped after 143.5 hr when the glucose concentration started to rise in the culture media indicating its lack of consumption. The strain reached a production 2.9 g/L β-carotene and a concomitant production of 40 g/L lipids and 31 g/L of citric acid (Supplementary Figure S6).

In order to further optimize bioreactor conditions we designed a culture media with double amounts of yeast extract and peptone, named Y20P40D, where the rate of glucose feeding was used as carbon source in fed-batch. In this experiment, a maximum production of 6.5 g/L of β-carotene was produced after 122 hr, the best titer described so far (Figure 8) and a yield of 0.036 g/g glucose. In this conditions, the production of citric acid was kept under 1 g/L along the fermentation while the lipid titer reached 40 g/L (Figure 8). Interestingly, the maximum β-carotene content was 89.6 mg/g of DCW, 50% higher than in flask and one of the highest described in the literature (Supplementary Table S3). Such results indicates that herein engineered *Y. lipolytica* using synthetic biology and metabolic engineering is a potential industrial producer of β-carotene.

4 | CONCLUSIONS

In this work we have combined traditional metabolic engineering strategies with novel synthetic biology tools in order to turn *Y. lipolytica* an industrially competitive producer of β-carotene. We found that increasing lipogenesis and gene copy number as well as using the most favorable set of promoters, greatly enhanced β-carotene production. Finally a fed-batch fermentation lead to the highest β-carotene production reported so far.

This work shows the enormous potential of *Y. lipolytica* to produce β-carotene in an economically feasible manner, not only by the high titer achieved but also due to the co-production of high amount of lipids, which can be used as fuels or chemicals (Singh et al., 2016). Moreover, the process can be further improved not only by strain engineering and bioreactor condition optimization but also by the use of low cost carbon sources such as starch or lignocellulosic materials. In addition, this work also highlights the importance of selecting the right microorganism and strain in a synthetic biology or metabolic engineering strategy. The selection of an oleaginous organism has permitted production levels far beyond the so far obtained in other yeasts such as *S. cerevisiae*, while the selection of a lipid overproducer strain further increased β-carotene content. This work is an example of how the rapid development of synthetic biology tools for DNA assembly and genome editing is facilitating the manipulation of non-conventional organisms, expanding the range of biotechnological chassis for metabolic engineering.

ORCID

Jean-Marc Nicaud http://orcid.org/0000-0002-6679-972X
Rodrigo Ledesma-Amaro http://orcid.org/0000-0003-2631-5898

REFERENCES

Awan, A. R., Blount, B. A., Bell, D. J., Shaw, W. M., Ho, J. C. H., McKiernan, R. M., & Ellis, T. (2017). Biosynthesis of the antibiotic nonribosomal peptide penicillin in baker's yeast. *Nature Communications*, 8, 15202.

Bailey, R., Madden, K., & Trueheart, J. (2012). Production of carotenoids in oleaginous yeast and fungi. US patent US8288149.

Barth, G., & Gaillardin, C. (1996). *Yarrowia lipolytica. Nonconventional yeasts in biotechnology: A handbook* (pp. 313–388). Berlin, Heidelberg: Springer Berlin Heidelberg.

Beekwilder, J., van Rossum, H. M., Koopman, F., Sonntag, F., Buchhaupt, M., ... Daran, J. M. (2014). Polycistronic expression of a beta-carotene biosynthetic pathway in Saccharomyces cerevisiae coupled to beta-ionone production. *Journal of Biotechnology*, 192(Pt B), 383–392.

Braunwald, T., Schwemmlein, L., Graeff-Honninger, S., French, W. T., Hernandez, R., Holmes, W. E., & Claupein, W. (2013). Effect of different C/N ratios on carotenoid and lipid production by Rhodotorula glutinis. *Applied Microbiology and Biotechnology*, 97(14), 6581–6588.

Celinska, E., Ledesma-Amaro, R., Larroude, M., Rossignol, T., Pauthenier, C., & Nicaud, J. M. (2017). Golden gate assembly system dedicated to complex pathway manipulation in Yarrowia lipolytica. *Microbial Biotechnology*, 10(2), 450–455.

Contreras, G., Barahona, S., Sepulveda, D., Baeza, M., Cifuentes, V., & Alcaino, J. (2015). Identification and analysis of metabolite production with biotechnological potential in Xanthophyllomyces dendrorhous isolates. *World Journal of Microbiology & Biotechnology*, 31(3), 517–526.

Dulermo, R., Brunel, F., Dulermo, T., Ledesma-Amaro, R., Vion, J., Trassaert, M., ... Leplat, C. (2017). Using a vector pool containing variable-strength promoters to optimize protein production in Yarrowia lipolytica. *Microbial Cell Factories*, 16(1), 31.

Dulermo, R., Gamboa-Melendez, H., Ledesma-Amaro, R., Thevenieau, F., & Nicaud, J. M. (2015). Unraveling fatty acid transport and activation mechanisms in Yarrowia lipolytica. *Biochimica Et Biophysica Acta*, 1851(9), 1202–1217.

Dulermo, R., Gamboa-Melendez, H., Ledesma-Amaro, R., Thevenieau, F., & Nicaud, J. M. (2016). Yarrowia lipolytica AAL genes are involved in peroxisomal fatty acid activation. *Biochimica Et Biophysica Acta*, 1861(7), 555–565.

Engler, C., & Marillonnet, S. (2013). Combinatorial DNA assembly using Golden Gate cloning. *Methods in Molecular Biology*, 1073, 141–156.

Engler, C., & Marillonnet, S. (2014). Golden gate cloning. *Methods in Molecular Biology*, 1116, 119–131.

Gao, S., Han, L., Zhu, L., Ge, M., Yang, S., Jiang, Y., & Chen, D. (2014). One-step integration of multiple genes into the oleaginous yeast Yarrowia lipolytica. *Biotechnology Letters*, 36(12), 2523–2528.

Gao, S., Tong, Y., Zhu, L., Ge, M., Zhang, Y., Chen, D., ... Yang, S. (2017). Iterative integration of multiple-copy pathway genes in Yarrowia lipolytica for heterologous beta-carotene production. *Metabolic Engineering*, 41, 192–201.

Grenfell-Lee, D., Zeller, S., Cardoso, R., & Pucaj, K. (2014). The safety of beta-carotene from Yarrowia lipolytica. *Food and Chemical Toxicology*, 65, 1–11.

Guo, Z., Duquesne, S., Bozonnet, S., Cioci, G., Nicaud, J. M., Marty, A., & O'Donohue, M. J. (2015). Development of cellobiose-degrading ability in Yarrowia lipolytica strain by overexpression of endogenous genes. *Biotechnology for Biofuels*, 8, 109.

Kerkhoven, E. J., Kim, Y. M., Wei, S., Nicora, C. D., Fillmore, T. L., Purvine, S. O., . . . Nielsen, J. (2017). Leucine biosynthesis is involved in regulating high lipid accumulation in Yarrowia lipolytica. *MBio*, 8(3), e00857–e00917.

Kerkhoven, E. J., Pomraning, K. R., Baker, S. E., & Nielsen, J. (2016). Regulation of amino-acid metabolism controls flux to lipid accumulation in Yarrowia lipolytica. *Npj Systems Biology and Applications*, 2, 16005.

Ledesma-Amaro, R., Dulermo, R., Niehus, X., & Nicaud, J. M. (2016). Combining metabolic engineering and process optimization to improve production and secretion of fatty acids. *Metabolic Engineering*, 38, 38–46.

Ledesma-Amaro, R., Dulermo, T., & Nicaud, J. M. (2015). Engineering Yarrowia lipolytica to produce biodiesel from raw starch. *Biotechnology for Biofuels*, 8, 148.

Ledesma-Amaro, R., Lazar, Z., Rakicka, M., Guo, Z., Fouchard, F., Coq, A. C., & Nicaud, J. M. (2016). Metabolic engineering of Yarrowia lipolytica to produce chemicals and fuels from xylose. *Metabolic Engineering*, 38, 115–124.

Ledesma-Amaro, R., & Nicaud, J. M. (2015). Yarrowia lipolytica as a biotechnological chassis to produce usual and unusual fatty acids. *Progress in Lipid Research*, 61, 40–50.

Ledesma-Amaro, R., & Nicaud, J. M. (2016). Metabolic engineering for expanding the substrate range of Yarrowia lipolytica. *Trends Biotechnology*, 34(10), 798–809.

Lin, Y., Jain, R., & Yan, Y. (2014). Microbial production of antioxidant food ingredients via metabolic engineering. *Current Opinion in Biotechnology*, 26, 71–78.

Liu, H. H., Ji, X. J., & Huang, H. (2015). Biotechnological applications of Yarrowia lipolytica: Past, present and future. *Biotechnology Advances*, 33(8), 1522–1546.

Madzak, C. (2015). Yarrowia lipolytica: Recent achievements in heterologous protein expression and pathway engineering. *Applied Microbiology and Biotechnology*, 99(11), 4559–4577.

Matthaus, F., Ketelhot, M., Gatter, M., & Barth, G. (2014). Production of lycopene in the non-carotenoid-producing yeast Yarrowia lipolytica. *Applied and Environmental Microbiology*, 80(5), 1660–1669.

Nam, H.-K., Choi, J.-G., Lee, J.-H., Kim, S.-W., & Oh, D.-K. (2013). Increase in the production of β-carotene in recombinant Escherichia coli cultured in a chemically defined medium supplemented with amino acids. *Biotechnology Letters*, 35(2), 265–271.

Nanou, K., & Roukas, T. (2016). Waste cooking oil: A new substrate for carotene production by Blakeslea trispora in submerged fermentation. *Bioresource Technology*, 203, 198–203.

Qiao, K., Wasylenko, T. M., Zhou, K., Xu, P., & Stephanopoulos, G. (2017). Lipid production in Yarrowia lipolytica is maximized by engineering cytosolic redox metabolism. *Nature Biotechnology*, 35(2), 173–177.

Reyes, L. H., Gomez, J. M., & Kao, K. C. (2014). Improving carotenoids production in yeast via adaptive laboratory evolution. *Metabolic Engineering*, 21, 26–33.

Sharpe, P., Ye, R., & Zhu, Q. (2014). Carotenoid production in a recombinant oleaginous yeast. US Patent US8846374.

Singh, G., Jawed, A., Paul, D., Bandyopadhyay, K. K., Kumari, A., & Haque, S. (2016). Concomitant production of lipids and carotenoids in rhodosporidium toruloides under osmotic stress using response surface methodology. *Frontiers in Microbiology*, 7, 1686.

Wasylenko, T. M., Ahn, W. S., & Stephanopoulos, G. (2015). The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in Yarrowia lipolytica. *Metabolic Engineering*, 30, 27–39.

Wei, H., Wang, W., Alahuhta, M., Vander Wall, T., Baker, J. O., Taylor, L. E., 2nd, . . . Zhang, M. (2014). Engineering towards a complete heterologous cellulase secretome in Yarrowia lipolytica reveals its potential for consolidated bioprocessing. *Biotechnology for Biofuels*, 7(1), 148.

Wichuk, K., Brynjolfsson, S., & Fu, W. (2014). Biotechnological production of value-added carotenoids from microalgae: Emerging technology and prospects. *Bioengineered*, 5(3), 204–208.

Xie, W., Ye, L., Lv, X., Xu, H., & Yu, H. (2015). Sequential control of biosynthetic pathways for balanced utilization of metabolic intermediates in Saccharomyces cerevisiae. *Metabolic Engineering*, 28, 8–18.

Yang, J., & Guo, L. (2014). Biosynthesis of β-carotene in engineered E. coli using the MEP and MVA pathways. *Microbial Cell Factories*, 13, 160.

Ye, V. M., & Bhatia, S. K. (2012). Pathway engineering strategies for production of beneficial carotenoids in microbial hosts. *Biotechnology Letters*, 34(8), 1405–1414.

Zhao, C. H., Cui, W., Liu, X. Y., Chi, Z. M., & Madzak, C. (2010). Expression of inulinase gene in the oleaginous yeast Yarrowia lipolytica and single cell oil production from inulin-containing materials. *Metabolic Engineering*, 12(6), 510–517.

Zhu, Q., & Jackson, E. N. (2015). Metabolic engineering of Yarrowia lipolytica for industrial applications. *Current Opinion in Biotechnology*, 36, 65–72.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

The developed GG method is now routinely used by all team members of the lab for the construction of different DNA assemblies and it is widely used across all the projects running within the team. Examples of different research axes of the team that have used the GG strategy developed here are summarized hereunder (Park et al., 2019; Soudier et al., 2019). The corresponding articles are available on the annexes of this manuscript (6.2). In a first work, UAS were identified in *Eyk* and *Eyd* promoter regions and were used to construct erythritol-inducible hybrid promoters. The promoters were design and constructed as bricks for the GG system, and assembled with a fluorescent protein (RedStarII) to analyse their expression capacity. It was shown that the promoter strenght increased with the number of tandem repeats of UAS. The new promoters enlarge the library of GG parts, and have practical applications in metabolic engineering and synthetic biology (Park et al., 2019). In a second example, we used the GG cloning strategy for the construction of multiple α-amylase-expression cassettes with the aim of identifying the best producer combination in *Y. lipolytica*. The *Sitophilus oryzae* alpha-amylase gene (SoAMY) was expressed under various promoters of different strength, and it was showed that the profile of the SoAMY activity differed significantly depending on the promoter used. Briefly, the hybrid promoter 8UASTEF is characterized by a higher activity compared to the native pTEF, and the erythritol induced promoter (pEYK1) was highly active during the first hours of the culture with a consequent reduced growth of the recombinant strain (Soudier et al., 2019). Both works had contributed to enlarge the GG library as well as on the characterisation of its parts. I was particularly involved in the desing of GG building blocks as well as in the GG assemblies.

In addition to the available toolkit and the proofs of a very performant system, a step by step detailed protocol to construct the library as well as the steps to assemble the building blocks has been submitted and is under review in the Methods in Molecular Biology book series. The expression of the carotenoid pathway was used as an example of the ability to assemble three heterologous genes together and express them in *Y. lipolytica* wild-type strains. This chapter is presented hereafter.

Title: Golden Gate multi-genes assembly method for *Yarrowia lipolytica*

Macarena Larroude[1], Jean-Marc Nicaud[1], Tristan Rossignol[1*]

[1] Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

* Corresponding author

Abstract

The oleaginous yeast *Yarrowia lipolytica* has emerged as a powerful alternative for biolipid production due to its high capacity for lipid accumulation. Genetic engineering and synthetic biology are promoted forward to improved production and reroute metabolism for high value compound synthesis. In this context, efficient, modular, and high-throughput compatible cloning and expression system are required to speed up and rationalize research in this field. Here we present the fast and modular Golden Gate cloning strategy for the construction of multigene expression vectors and their transformation into *Yarrowia lipolytica*. As an example, we used the heterologous expression of the carotenoid pathway by cloning three genes involved in this pathway in only one vector allowing reaching production of β-carotene after a single transformation.

Key words: *Yarrowia lipolytica*, Golden Gate, gene assembly, synthetic biology, gene expression, β-carotene production

3.1.2   **CRISPR**

Clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-as-sociated (Cas) (altogether CRISPR-Cas) are originally adaptive immune systems of bacterial and archaeal providing them protection from viruses and plasmids. Briefly, fragments of the invading DNA are incorporated, as 'spacers', into CRISPR loci, short CRISPR RNAs (crRNAs) are then synthesized and used to guide Cas nucleases to the homologous foreign DNA for degradation (Jinek et al., 2012) (Figure 16).
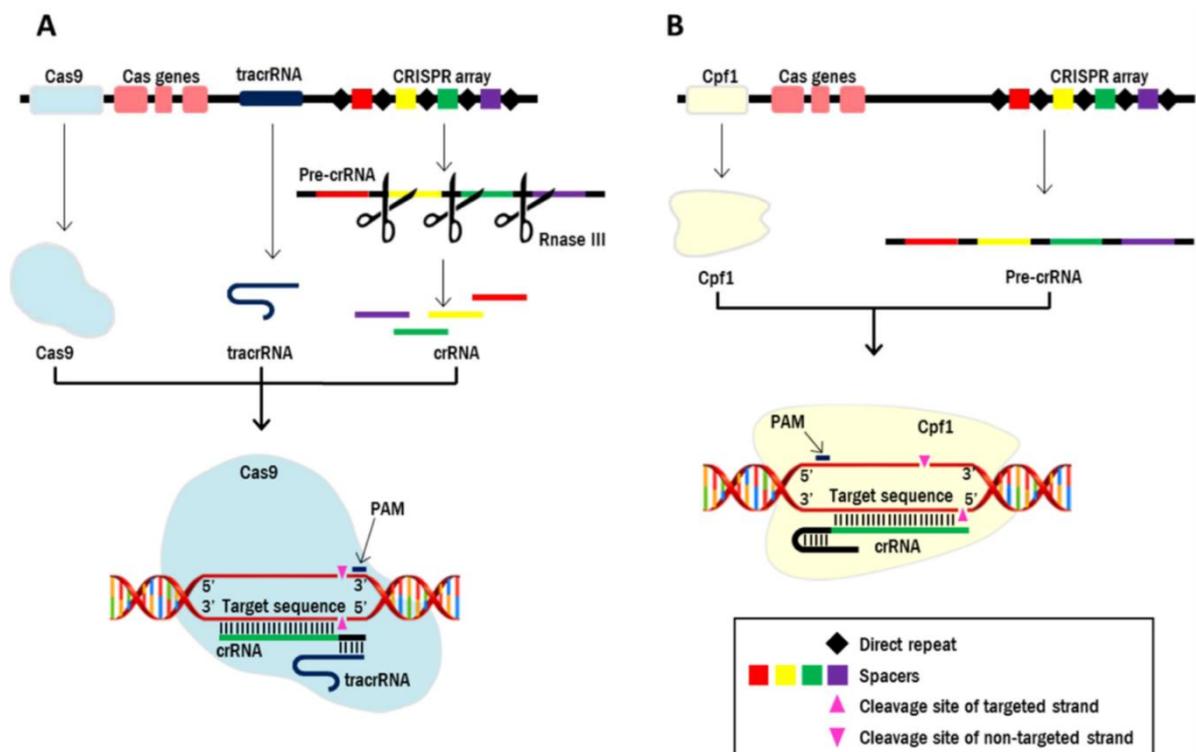
Swiat *et al.* 2017

Figure 16. Schematic representation of Cas9 and Cpf1 native systems.

Each CRISPR locus encodes acquired 'spacers' that are separated by repeat sequences. Transcription of the locus yields a pre-crRNA, which is then processed into crRNA that guide effector nuclease complexes to cleave dsDNA sequences complementary to the spacer ('protospacer'). Compared to Cas9, Cpf1 does not require a tracrRNA, has a T-rich PAM sequence located at the 5' end of the protospacer, is capable to mature its own crRNA array, cleaves DNA distal from the PAM and generates staggered ends. Cas genes are enzymes involved in the incorporation of foreign DNA into the CRISPR loci in the native system where the CRISPR-Cas9 protects prokaryotes from invading DNA. Figure from Swiat et al., 2017

Highly diverse Cas proteins are involved in CRISPR systems, defining different systems which classification was recently updated by Koonin and co-workers (Koonin et al., 2017; Makarova et al., 2015). Some CRISPR-Cas systems have been engineered to perform robust RNA-guided genome modifications in multiple cell types, from yeast

to human cells. The generated systems are easy to use and highly efficient, and substantially improved the ease of genome editing (Mali et al., 2013).

The eukaryotic setting involves expression of a codon-optimized Cas protein with an appropriate nuclear localization signal (NLS) to address the enzyme into the nucleus, and the RNA structure components, expressed via a RNA polymerase promoter, that guide the enzyme to the DNA site to be edited.

Among Cas enzymes, Cas9 has been the most extensively used, specially the one from *Streptococcus pyogenes*, also known as SpCas9. To be guided towards the editing site, this endonuclease requires two RNA structures: (i) the crRNA harboring a 20-nt sequence, homologous to the target sequence, that guides the endonuclease enzyme towards the editing site, and (ii) a trans-activating RNA (tracrRNA) that binds to the crRNA and to the SpCas9. To simplify the heterologous expression of the RNA molecule and increase the efficiency of the system, a chimeric single guide RNA (sgRNA) containing crRNA and tracrRNA already fused has been constructed. The sgRNA can be expressed from its own promoter, avoiding the requirement of an RNase for processing the precursor RNA transcript (Figure 16). As all known CRISPR associated endonucleases, SpCas9 can only cut DNA located near a PAM (protospacer-adjacent motif) sequence. The SpCas9 PAM sequence is G-rich, 5'-NGG-3' (being N any nucleotide), and it is located at the 3'end of the target (protospacer) sequence (Jinek et al., 2012). At the target site, Cas9 cuts the double strand of DNA which can be repaired by NHEJ, leading to indels that produce frameshift mutation and gene knockouts, or by HR and integrate heterologous DNA sequences. Cas9 has also been engineered as a nuclease-deficient enzyme (dCas9), with the capacity to bind the DNA but unable to cut it. This dCas9 can be fused to different functional proteins or effector domains, greatly extending the CRISPR toolbox for a wide range of applications exemplified by inhibition or activation of endogenous gene expression, epigenetic modification, and gene visualization (Mali et al., 2013). Cas9 is not the only enzyme engineered, nowadays Cpf1 and Cas13 are also being used, increasing the type of sequences that can be targeted and the RNA processing methods (Freije et al., 2019; Swiat et al., 2017; Zetsche et al., 2015). One can expect that these functionalities will continue to increase as there is a broad spectrum of nucleases that has not been fully exploited.

In *Y. lipolytica* genetic modifications are challenging. The available genome editing tools are limited by low-efficiency HR, and sequential integrations of expression cassettes require a selectable marker that has to be afterwards recovered by Cre/Lox system. The emergence of the CRISPR-Cas system offered a potential solution due to its high efficiency, ease of operation, site-specific modification and potential marker free editing.

As previously mentioned, CRISPR-Cas system was successfully implemented in *Y. lipolytica* for the first time by Schwartz *et al.* (Schwartz et al., 2016) and rapidly expanded afterwards, allowing gene knock-out/knock-in and repression/activation applications which are very helpful for accelerating engineering cycles. The different approaches developed for *Y. lipolytica* were described in the introduction section 1.3.

During the present work, a CRISPR-Cas system was set up with the aim of having a versatile and modular system. Taking advantage of the available, ready to use building blocks library, a GG strategy for the construction of CRIPSR-Cas vectors was designed. The developed system eases the construction of vectors with different markers and/or different Cas9 promoters, and allows the easy and rapid introduction of any gRNA sequence. The method was designed in a two-step cloning approach (Figure 17). During the first step, the platform vector is constructed based on BsaI restriction enzyme and reusing GG building blocks (promoter, marker and bacterial component). This platform vector contains: (1) a bacterial component (pBS1A3) for selection (Ampicillin resistance gene) and propagation (ColE1 region) in *E. coli*, (2) the sgRNA platform containing the *E. coli* chromophore RFP gene flanked by BsmBI sites; (3) the *Y. lipolytica* marker; (4) the *Y. lipolytica* centromeric region ylARS1 for propagation and replication of the vector in the yeast; (5) the Cas9 promoter; and (6) the Cas9 fused to the NLS (SV40) and CYCt as terminator. The second step uses a different IIS enzyme, BsmbI, for the cloning of gRNA. The platform vector harbors a reporter system, the bacterial chromophore RFP, flanked by BsmBI recognition sites for positive clones screening. The digestion of the platform vector (acceptor vector) with this enzyme release the RFP gene and allow the gRNA to be inserted at the position. After *E. coli* transformation, only white clones are screened for the correct assembly, similarly to the original GG

method strategy. A vector assembly strategy as well as a detailed protocol were designed and are presented in section 2.8.
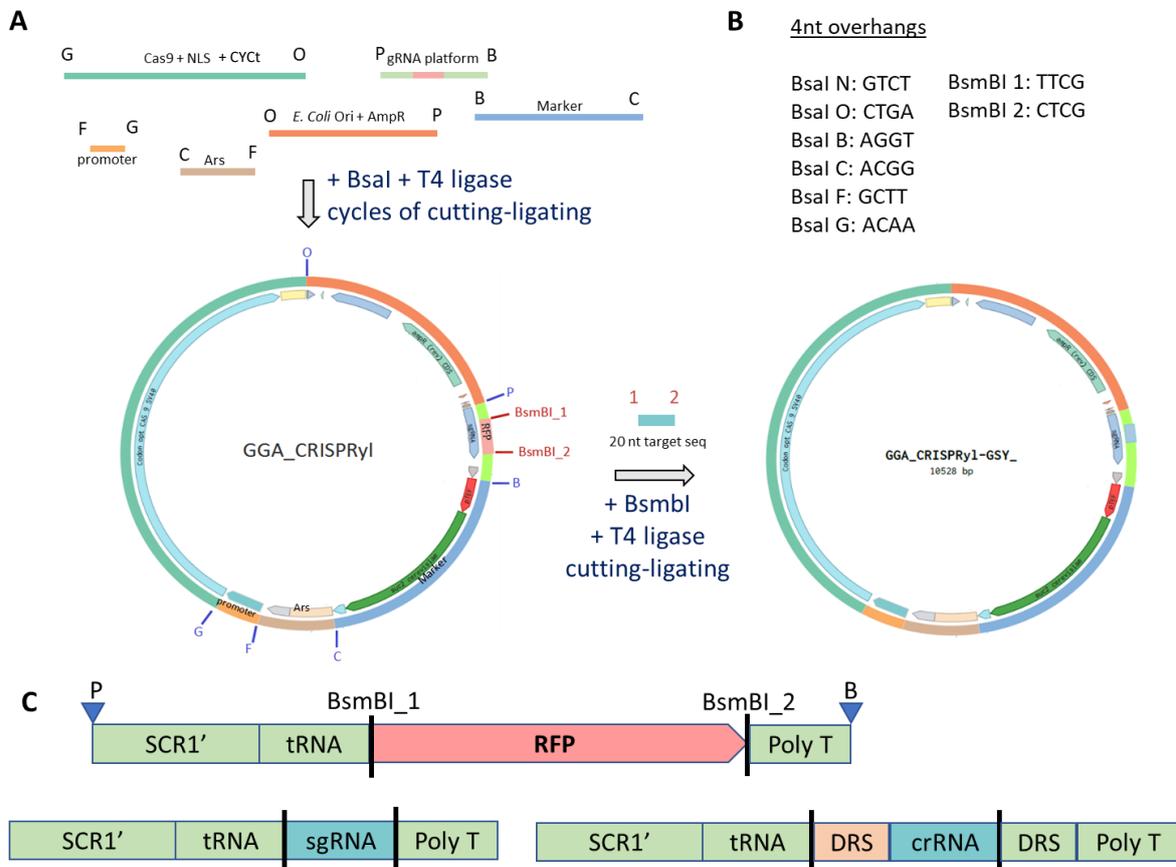
Figure 17. Modular CRISPR strategy based on GG assembly.

A- Schematic representation of the modular two-step cloning strategy for the construction of CRISPR-Cas vector. In a first step, the building blocks are assembled to construct a platform vector containing a chromophore (RFP) at the place of the gRNA. In the second step, the RFP is released by BsmBI and the fragment with the 20nt specific to the gRNA are incorporated. Thus, only white clones are screen after cloning. **B-** The 4nt overhangs specific for each DNA fragment are listed. It is also specified the enzyme that is used to release the fragment from the donor vector. **C-** Representation of the structure of the DNA block harbouring the RFP (gRNA platform). It shows the places where BsmBI cuts, with the consequent release of the RFP. The sites P and B correspond to the 4nt overhangs produced after BsaI digestion and used to construct the platform vector. In the bottom part of the figure, the obtained structures after the cloning of the gRNA-specific sequence are represented. The left part corresponds to the Cas9 gRNA and the right to the Cpf1 gRNA. The gRNA and the endonuclease are the only differences between these two systems. RFP: Red Fluorescent Protein.

As a result of this work, a set of platform vectors was constructed (Table 2), validated for their genome editing capacity and efficiency in *Y. lipolytica* and is available via Addgene (Nos 129656-129661) (https://www.addgene.org/). Each vector contains a different selection marker, thus allowing genome editing in different genetic background strains including wild-type strains, and are ready to add any gRNA of interest.

Table 2. List of platform CRISPR-Cas9 vectors.

GGA-Cas9-ylLEU2-gRNA::RFP	Leucine selection
GGA-Cas9-ylURA3-gRNA::RFP	Uracyl selection
GGA-Cas9-ylLYS5-gRNA::RFP	Lysine selection
GGA-Cas9-SUC2-gRNA::RFP	Sucrose selection
GGA-Cas9-HYG-gRNA::RFP	Hygromycin resistance
GGA-Cas9-NAT-gRNA::RFP	Nourseothricin resistance

Each vector has a different selection marker and is ready for gRNA cloning. All of them have 8UAS1TEF as Cas9 promoter and are available through Addgene. RFP: Red Fluorescent Protein, a bacterial chromophore for rapid positive clone screening.

The use of these CRISPR-Cas9 vectors strongly improved the HR at the cutting site when a homologous DNA template donor is simultaneously delivered in *Y. lipolytica*. More than 80% of successful integration of a cassette at a specific locus was reached, compared with only 15-20% of HR obtained when no CRISPR system is used. Moreover, the system also allowed to drastically reduce the length of the homologous recombination flanking region for cassette integration, from 1000 to 100 bp, without having to use a Δ*ku70* background strain and without compromising the correct integration rate. The 100 bp sequence achieved up to 25% of successful integration, and 88% in Δ*ku70* strains. In addition, the modification of more than one genomic site at the same time was also demonstrated. For this, the yeast was simultaneously transformed with CRISPR-Cas9 plasmids targeting *URA3* and *GSY* genes. 50% of double disruption was reached. However, the efficiency depended on the gRNA and the selection marker used in each plasmid.

Several genes were targeted with the Cas9 system, *URA3*, *EYK*1, *EYD*1, which can be used as markers; and *GSY*1, *MFE2* and *LIP*2, related to lipid metabolism. It was evidenced that the edition rate depends on the targeted gene and on the sgRNA used. These results are in accordance with what was observed in other works (Mans et al., 2015; Schwartz et al., 2019). The test of these vectors also showed that the combination strain-marker had an influence on the editing rate. In addition, these vectors can be employed to knock out auxotrophic or "catabolic selectable markers" (CSM) genes (like *URA3* or *EYK*) and thus increase the panel of markers available in any strain.

The results of this work are detailed in the article published in Biotechnology Letters and are presented below (Larroude et al., 2020).

ORIGINAL RESEARCH PAPER

A set of *Yarrowia lipolytica* CRISPR/Cas9 vectors for exploiting wild-type strain diversity

Macarena Larroude · Heykel Trabelsi · Jean-Marc Nicaud · Tristan Rossignol

Abstract

Objectives The construction and validation of a set of *Yarrowia lipolytica* CRISPR/Cas9 vectors containing six different markers that allows virtually any genetic background to be edited, including those of wild-type strains.

Results Using the Golden Gate method, we assembled a set of six CRISPR/Cas9 vectors, each containing a different selection marker, to be used for editing the genome of the industrial yeast *Y. lipolytica*. This vector set is available via Addgene. Any guide RNA (gRNA) sequence can be easily and rapidly introduced in any of these vectors using Golden Gate assembly. We successfully edited six different genes in a variety of genetic backgrounds, including those of wild-type strains, with five of the six vectors. Use of these vectors strongly improved homologous recombination and cassette integration at a specific locus.

Conclusions We have created a versatile and modular set of CRISPR/Cas9 vectors that will allow any *Y. lipolytica* strain to be rapidly edited; this tool will facilitate experimentation with any prototroph wild-type strains displaying interesting features.

Keywords CRISPR/Cas9 · *Yarrowia lipolytica* · Golden Gate · GSY1 · Synthetic biology

M. Larroude · H. Trabelsi ·
J.-M. Nicaud · T. Rossignol (✉)
Université Paris-Saclay, INRAE, Micalis Institute,
78350 Jouy-en-Josas, AgroParisTech, France
e-mail: tristan.rossignol@inrae.fr

Introduction

Yarrowia lipolytica is widely used as a microbial cell factory chassis in the development of industrial applications aiming to produce fatty acids, organic acids, or enzymes (Nicaud 2012; Madzak 2015; Ledesma-Amaro and Nicaud 2016). Although many engineering tools are now available (Larroude et al. 2018), the high rate of non-homologous end joining (NHEJ) impairs the efficiency of targeted genome modification. Moreover, limitations related to selection markers and the need to recycle them make engineering efforts even more time-consuming, and it can be challenging to generate highly modified strains with traits that correspond to the demands of industrial processes. CRISPR/Cas9 technology is being continually refined, and it was rapidly implemented in *Saccharomyces cerevisiae* (DiCarlo et al. 2013). It has also been successfully used in *Y. lipolytica*, with the aim of overcoming the aforementioned limitations and accelerating engineering cycles.

Several CRISPR/Cas9 systems dedicated to *Y. lipolytica* have recently been described. Both an RNA polymerase II (Pol II) transcription system and RNA polymerase III (Pol III) elements have been set up. Gao et al. (Gao et al. 2016) developed a Pol II gene disruption system that has successfully yielded single to triple mutations in a single step. Schwartz et al. (Schwartz et al. 2016) used a Pol III system and developed an efficient synthetic Pol III-tRNA hybrid promoter for gene disruption. In addition, a guide RNA (gRNA) system involving the expression of orthogonal T7 polymerase was exported to *Y. lipolytica*; it was based on the pCRISPRyl vector of Schwartz et al. (Morse et al. 2018). Structure–function relationships study and methodological research focused on single point mutations has also been described (Borsenberger et al. 2018). In addition, alternative systems, such as using defective Cas9 (CRISPRi) to inhibit expression, CRISPRa to activate gene expression, and the dual CRISPR/Cas9 system to excise and integrate genes have been shown to function in *Y. lipolytica* (Schwartz et al. 2017a, 2018; Gao et al. 2018).

It is now well established that the CRISPR/Cas9 system functions properly in *Y. lipolytica*, and modular and robust tools are needed for this system to become the standard for editing the genome of this yeast species. Such tools are particularly lacking when it comes to engineering the diversity of wild-type strains. These strains are never used as there are no markers available for them, which hinders metabolic engineering. To date, researchers have mainly concentrated on "laboratory" strains for which auxotrophic markers are already available. However, it is now well known that wild-type isolates display a wide range of traits when it comes to producing lipids, citric acids, or polyols, for example (Egermeier et al. 2017; Quarterman et al. 2017; Carsanba et al. 2019); in these experiments, some of the isolates strongly outperformed the standard laboratory strains. Wild-type strains may also respond very differently to environmental conditions and control parameters (Egermeier et al. 2017).

Having selectable markers is crucial to the success of genome editing technologies. In *Y. lipolytica*, such markers were initially based on leucine (LEU2), uracil (URA3), lysine (LYS5) and adenine (ADE2) auxotrophies (Barth and Gaillardin 1996). The first dominant markers to be developed relied on the expression of the *E. coli hph* gene, which confers antibiotic resistance to hygromycin B (Barth and Gaillardin 1996). Additional

markers have been developed more recently. They are based on the *Streptomyces noursei Nat1* gene (which provides resistance to nourseothricin) (Kretzschmar et al. 2013); the *Y. lipolytica AHAS* gene (which provides resistance to the herbicide chlorimuron ethyl); the *E. coli guaB* gene (which provides resistance to mycophenolic acid) (Wagner et al. 2018); the *Streptoalloteichus hindustanus ble* gene (which provides resistance to zeocin) (Tsakraklides et al. 2018); and the phosphite dehydrogenase *ptxD* gene from *Pseudomonas stutzeri* (which allows *Y. lipolytica* to grow on potassium phosphite in phosphate-deficient media) (Shaw et al. 2016). Additional markers have been developed that are related to the ability to catabolize carbon sources (hereafter referred to as "catabolic selectable markers" [CSM]). These markers have the advantage of not being involved in essential metabolic pathways. For instance, the *S. cerevisiae SUC2* gene, which encodes invertase, has been used to select transformants on sucrose media (Nicaud et al. 1989). More recently, a novel CSM was developed that is centered on the *EYK1* gene, which encodes an erythrulose kinase (Vandermies et al. 2017). This enzyme participates in an early step in erythritol catabolism and is essential for cell growth on erythritol-based medium (Carly et al. 2017).

Here, we describe the construction and validation of a set of seven CRISPR/Cas9 vectors, which each contain a different selection marker. They allow genome editing within virtually any genetic background and are particularly useful for working with wild-type strains, thanks to the panel of dominant markers they represent. Having access to such tools is especially important when it is necessary to integrate multiple cassettes to generate large pathways, which involves the use of multi-auxotrophic strains. gRNA sequences can easily be cloned into these vectors via the Golden Gate method. Consequently, these replication-based CRISPR/Cas9 vectors can be employed to knock out auxotrophic or CSM genes and thus increase the panel of markers available in any strain.

Materials and methods

Strains and media

The *Escherichia coli* and *Y. lipolytica* strains and plasmids used in this study are listed in Table S1.

E. coli strain DH5α was used for cloning and plasmid propagation. The transformation of chemically competent *E. coli* cells was performed using a heat shock protocol. Cells were grown at 37 °C with constant shaking on 5 ml of LB medium (10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl) that contained ampicillin (100 µg/ml) or kanamycin (50 µg/ml) for plasmid selection. For yeast selection and growth, minimal YNBD medium containing 10 g/L glucose (Sigma), 1.7 g/L yeast nitrogen base (YNBww; Difco), 5.0 g/L NH$_4$Cl, and 50 mM phosphate buffer (pH6.8) was used. To meet auxotrophic requirements, uracil (0.1 g/L), lysine (0.8 g/L), and leucine (0.1 g/L) were added to the culture medium when necessary. Erythritol and lysine were used as carbon sources at concentrations of 10 g/L and 0.1 g/L, respectively, to allow the selection of *Δeyk1* and *Δeyd1* mutants; lysine was used as an additional carbon source to boost growth in the absence of glucose, but in a concentration not sufficient for *Δeyk1* and *Δeyd1* to grow in presence of erythritol and absence of glucose. An oleic acid emulsion (0.05%) was used as a carbon source for *Δmfe* selection. Tributyrin YNB medium was utilized for *Δlip2* selection as described in (Pignede et al. 2000). For antibiotic selection, hygromycin (250 µg/ml) or nourseothricin (400 µg/ml) was added to rich YPD medium (20 g/L Bacto Peptone, 10 g/L yeast extract, and 20 g/L glucose). Solid media were prepared by adding 15 g/L agar (Invitrogen) to liquid media.

Construction of acceptor vectors for gRNA cloning

Six fragments were used to assemble the CRISPR/Cas9 acceptor vectors. Fragment 1 was the pSB1A3 bacterial plasmid containing the ampicillin resistance gene and the ColE1 region for selection and propagation in *E. coli*. Fragment 2 was the sgRNA module with the BsmBI recognition sites flanking the red fluorescent protein (RFP) gene, which was used as a chromophore in *E. coli* and which replaced the 20-nt target sequence (Fig. 1b). Fragment 3 was the excisable marker flanked by the I-sceI sites and the LoxP/LoxR. Fragment 4 was a 227-bp portion of the ARS68 sequence described in (Fournier et al. 1991), which allows plasmid replication and segregation in *Y. lipolytica*. Another 105-bp portion of the ARS68 sequence was present at the 3′ end of the sgRNA
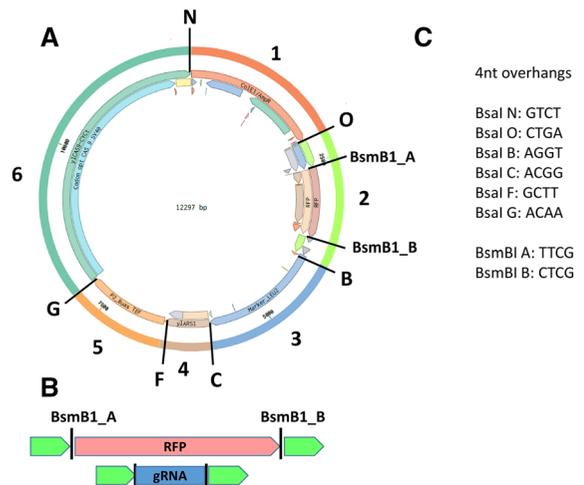
Fig. 1 Schematic draw showing an example of a CRISPR/Cas9 acceptor vector assembled via the Golden Gate method. **a** Assembly of the six fragments: the *E. coli* bacterial component (fragment 1); the sgRNA region containing the *E. coli* red fluorescent protein chromophore (RFP) gene flanked by the BsmBI sites (fragment 2); the excisable *Y. lipolytica* marker ylLEU2ex (fragment 3); the *Y. lipolytica* centromeric region ylARS1 (fragment 4); the promoter region P2_8UAS TEF (fragment 5); and the Cas9-SV40 region (fragment 6). **b** Cutting with BsmBI releases the RFP gene and permits the assembly of the 20-nt target sequence at that same location. **c** BsaI and BsmBI overhang sequences

module and was thus cloned into the vector as part of fragment 2. Fragment 5 was the strong hybrid promoter p8UAS1TEF described by (Celinska et al. 2017), which is used in the expression of the endonuclease. Fragment 6 was the codon-optimized *Streptococcus pyogenes* Cas9-SV40 system including the CYC-t terminator described by Schwartz et al. (Schwartz et al. 2016). The internal BsmBI and BsaI sites were removed from all the fragments, which were then synthetized by Twist Bioscience or GeneMill.

To build fragment 2, the fragment carrying the RFP gene was amplified using primer pair RFP-SfiI-Fw/RFP-SfiI-Rv, and the plasmid GGE029 was employed as a template (Supplementary Table S1). Then, the RFP gene flanked by the SfiI sites was inserted into the sgRNA module (JME4315), which itself included the same SfiI site, giving rise to sgRNA::RFP JME4366 (Table 1).

The acceptor vectors were assembled using the Golden Gate method in accordance with the Larroude et al. protocol (Larroude et al. 2019). Plasmid DNA was extracted using a commercial miniprep kit (NucleoSpin® Plasmid, Macherey–Nagel) in

Table 1 Golden Gate fragments available for CRISPR/Cas9 vector assembly

Fragment #	E. coli/plasmid name	Region name	Region description	5′ Bsa1	3′ Bsa1	Reference/source
1	GGE124	AmpR-ORI1	Bacterial vector	GTCT	CTGA	(Celinska et al. 2017)
2	JME4366	sgRNA::RFP	sgRNAplatform site	CTGA	AGGT	This study
3a	GGE105	ylLEU2ex	Excisable *LEU2 Y. lipolytica* marker	AGGT	ACGG	(Celinska et al. 2017)
4	JME4313	ylARS1	Replication in *Y. lipolytica*	ACGG	GCTT	This study
5	GGE152	P2_8UAS TEF	Promoter for Cas9 expression	GCTT	ACAA	(Celinska et al. 2017)
6	JME4311	Cas9-SV40	Codon-optimized Cas9-SV40	ACAA	GTCT	This study
3b	GGE216	ylURA3ex	Excisable *URA3 Y. lipolytica* marker	AGGT	ACGG	(Fickers et al. 2003)
3c	GGE176	ylLYS5ex	Excisable *LYS5 Y. lipolytica* marker	AGGT	ACGG	(Celinska et al. 2017)
3d	GGE367	HPHex	Hygromycin B resistance	AGGT	ACGG	(Larroude et al. 2019)
3e	GGE368	NATex	Nourseothricin resistance	AGGT	ACGG	(Larroude et al. 2019)
3f	GGE268	EYK1ex	Growth on erythritol	AGGT	ACGG	This study

accordance with the manufacturer's instructions. Correct assembly was verified by colony PCR and by plasmid digestion with the BglII restriction enzyme. To build the different acceptor vectors, alternative versions of fragment 3 with the different *Y. lipolytica* markers were employed (from 3a to 3f, see Table 1).

Digestion of the acceptor vectors with BsmBI released the RFP gene and allowed the 20-nt target sequence to be inserted at that position. This process made it possible to rapidly visually screen for positive *E. coli* clones that had correctly assembled the CRISPR/Cas9 vectors, in which the gRNA had replaced the RFP gene.

gRNA design and plasmid construction

The gRNA was designed with the CRISPOR tool (https://crispor.tefor.net/). We chose target sequences with high efficiency scores and low numbers of predicted off-target sites; our preference was for target sequences in the middle of the open reading frame. The target sequences were 20 nt in length and did not include the PAM sequence. The gRNA was introduced into the acceptor vectors by annealing two overlapping oligonucleotides that generated overhangs matching those of the BsmBI site of the acceptor vectors. The oligonucleotides had the following structure: forward oligonucleotide: 5′TTCGATTCCGGGTCGGCGCA GGTTGNNNNNNNNNNNNNNNNNNNNGTTTT- A 3′ reverse oligonucleotide: 5′GCTCTAAAACNNN NNNNNNNNNNNNNNNNNNCAACCTGCGCCGA

CCCGGAAT 3′, where N represents the 20-nt target sequence. The underlined letters indicate the 4-nt overhang. The complementary gRNA oligonucleotides were phosphorylated and annealed as follows: we mixed 1 μL T4 Kinase (New England Biolabs, Ipswich, MA), 1 μL forward oligonucleotide (100 μM), 1 μL reverse oligonucleotide (100 μM), 1 μL T4 Ligase Buffer (10X; New England Biolabs), and 6 μL H$_2$O; we incubated the mixture at 37 °C for 30 min and then heated the mixture so it stayed at 95 °C for 5 min; and we allowed the temperature to drop back down to 25 °C at a rate of 5 °C min^{-1}. The reaction mixture was diluted with water (ratio of 1:200) before being assembled with the CRISPR acceptor vector.

The gRNA double-stranded insert was assembled with the pGGA_CRISPRyl acceptor vector via the Golden Gate method as follows: we mixed 2 μL gRNA insert, 100 ng pGGA_CRISPRyl, 2 μL T4 Ligase Buffer (New England Biolabs), 1 μL BsmBI (New England Biolabs), 1 μL T7 Ligase (New England Biolabs), and 20 μl H$_2$O, and we incubated the mixture in a thermocycler using the following program: [5 min at 55 °C, 5 min at 16 °C] × 30, 5 min at 50 °C, and 5 min at 80 °C. We used 10 μL of the assembly mix to transform *E. coli* and then grew the bacteria on LB ampicillin plates. Plasmid DNA was extracted using a commercial miniprep kit (NucleoSpin® Plasmid, Macherey–Nagel) in accordance with the manufacturer's instructions. Correct assembly of the positive transformants (white *E. coli*) was verified either by colony PCR or by plasmid digestion

with the BglII restriction enzyme. The status of positive clones was confirmed via sequencing.

Transformation of the pGGA_CRISPRyl-gRNA plasmids to perform gene editing in *Y. lipolytica*

Transformation of *Y. lipolytica* was performed using the lithium-acetate method (Barth and Gaillardin 1996). Cells were either plated directly onto selective media for the direct selection of transformants (no outgrowth step), or an outgrowth step was performed on selective liquid media and allowed recovery. In the outgrowth step, cells were cultured in 5 mL of selective media for 24 h and then diluted to obtain 50–100 colonies per plate of rich YPD medium. The transformants could then be tested for the desired phenotype. The gene disruption success rate was determined by counting the number of colonies that showed the expected phenotype; the status of positive clones was confirmed via sequencing. After the screening step, the clones with the desired phenotype were grown in rich liquid YPD medium for 12–24 h to cure the CRISPR\Cas9 plasmid.

Transformation using the pGGA_CRISPRyl-gRNA plasmids and the deletion cassette for specific locus integration

The deletion cassette was composed of a *URA3*ex expression cassette flanked upstream and downstream by the homologous recombination sequences for the *GSY1* gene. Overall structure of the cassette and construction has been described elsewhere (Fickers et al. 2003). *Y. lipolytica* was transformed using the lithium-acetate method (Barth and Gaillardin 1996). Then, 500 ng of the CRISPR-Cas9 vector was co-transformed with 500 ng of the NotI-digested deletion cassette. The transformation reaction was inoculated in 9 mL of selective liquid medium and cultured for 48 h. One ml of the culture was then transferred into YPD medium for 24 h to allow plasmid curing. Finally, the culture was diluted and plated so as to obtain 50–100 colonies per plate.

When the deletion cassette was transformed without the CRISPR\Cas9 vector, no outgrowth was performed, and 200 μL of the transformation reaction was directly grown on plates containing selective agar.

Mutant phenotype analysis

After the transformation with the CRISPR\Cas9 vector, the outgrowth step, and the plating, the colonies could undergo phenotype screening.

gsy1 disruption

When *gsy1* had been successfully disrupted, the phenotype was an absence of glycogen accumulation, which could be visualized using staining with Lugol's iodine (prepared by mixing water solutions of 2% KI and 1% I_2 at a ratio of 1:1). Colonies in which *gsy1* had been successfully disrupted remained clear, while colonies containing active GSY1 were brown. There are two screening options. The clones can be tested on their plates, by adding 4 mL of Lugol's iodine. They can also be tested in 96-well microplates after 24 h of culture on YPD medium: 30 μL of Lugol's iodine is added to each well after the supernatant culture has been eliminated.

eyk1 and *eyd1* disruption

Transformants were grown on both YPD and YNB-erythritol-lysine media for 48 h at 28 °C. Then, we assessed whether clones were able to use erythritol as a carbon source. When a clone grew on YPD but not on YNB-erythritol-lysine, it indicated that *eyk1* or *eyd1* had been disrupted. The JMY7126 strain (*Δeyk1*) was used as the positive control.

mfe1 disruption

Clones in which *mfe1* had been disrupted should be unable to use oleic acid as a carbon source. Transformants were grown on both YPD and YNB-oleic acid media for 48 h at 28 °C. When a clone grew on YPD medium but not on YNB-oleic acid medium, it indicated that *mfe1* had been disrupted. The strain JMY1888 (*Δmfe1*) was used as a positive control.

ura3 disruption

Transformants were grown on both YPD medium and YNB medium without uracil for 48 h at 28 °C. When a clone grew on YPD but not on YNB without uracil, it indicated that *ura3* had been disrupted.

lip2 disruption

The *LIP2* gene encodes the main extracellular lipase, so Δ*lip2* strains display a reduced halo of triglyceride hydrolysis on YNB medium containing tributyrin (Pignede et al. 2000). Transformants were grown on YNB medium containing tributyrin for 48 h at 28 °C. Clones with strongly reduced halos were considered to carry a disrupted version of *LIP2*.

Colony PCR and sequence-based verification of gene disruption

For positive clones, we verified gene disruption via sequencing. We employed specific primers that spanned the gRNA target regions (supplementary Table S2). First, the regions were amplified by colony PCR. Single colonies were picked with a tip and transferred to 2μL water and cells were lysed with a 15 min heating cycle in the thermocycler. The cells lysed were directly used in the standard PCR reaction. The amplified fragments were then purified using gel extraction and PCR purification kits (NucleoSpin® Gel and PCR Clean-Up, Macherey–Nagel) and sequenced. The sequences were aligned against reference sequences to identify the mutations in the target sequence.

Results

Construction of a versatile CRISPR/Cas9 vector set

To extend the set of synthetic biology tools available for *Y. lipolytica*, we wished to develop a systematic backbone method for building a set of CRISPR/Cas9 vectors. Our design was based on the basal structure of the plasmid developed by Schwartz et al. (Schwartz et al. 2016), and the Golden Gate method made it possible to assemble and switch out different parts. We took advantage of a large set of Golden Gate bricks that were recently made available for *Y. lipolytica* (Larroude et al. 2019). For our specific CRISPR/Cas9 vector set, the necessary parts include i) a bacterial plasmid containing the ampicillin resistance gene and the ColE1 region for selection and propagation in *E. coli*; ii) a *Y. lipolytica* selection marker; iii) the ARS/CEN fragment allowing plasmid replication in *Y.*

lipolytica; iv) a *Y. lipolytica* promoter for endonuclease expression; v) the Cas9 endonuclease optimized for *Y. lipolytica*; and vi) a region for cloning gRNA. In association with the latter, we implemented a direct screening method for verifying that the recombinant plasmid had incorporated the gRNA: we integrated a RFP chromophore gene at the gRNA cloning site that would be later released, allowing the assembly of the 20-nt target sequence at that same location (Celinska et al. 2017). The set of bricks designed and used to assemble the CRISPR/Cas9 acceptor vectors are listed in Table 1. These acceptor vectors were first put together using BsaI overhang sequences, as described in Larroude et al. (Larroude et al. 2019). These acceptor vectors were then ready to be assembled with the gRNA fragment of choice using the BsmBI overhang sequences, as described in the Material and Methods. The overall design is depicted in Fig. 1; the Golden Gate overhang sequences, the BsaI overhang sequences for backbone construction, and the BsmBI overhang sequences for gRNA cloning are indicated.

Using the pool of Golden Gate bricks, we built a set of acceptor vectors that were ready for immediate use in gRNA cloning employing different markers. We used LEU2ex, URA3ex, and LYS5ex for the auxotrophic markers; NATex and HPHex for the antibiotic markers; and EYK1ex for the CSM markers. All the acceptor vectors are listed in Table 2.

Validation of genome editing

We tested our CRISPR/Cas9 system by first targeting genes that can be used as markers: *EYK1*, *EYD1*, and *URA3*. The *EYK1* gene encodes erythrulose kinase, which is involved in erythritol catabolism. The disruption of this gene allows the selection of strains that cannot use erythritol as their sole carbon source (Carly et al. 2017). The *EYD1* gene encodes erythritol dehydrogenase, which is also involved in erythritol catabolism (Carly et al. 2018). The *URA3* gene is involved in uracil metabolism. The disruption of *EYK1* and *EYD1* increased the strength of recently developed erythritol-inducible promoters pEYK1 and pEYD1 (Trassaert et al. 2017; Park et al. 2019). It is therefore particularly useful to disrupt these genes in strains that contain expression cassettes based on these promoters.

We then extended our testing to include genes related to lipid metabolism: *MFE2*, *GSY1*, and *LIP2*. The *MFE2* gene encodes the multifunctional enzyme,

Table 2 List of the gRNA acceptor vectors that were built

Strain number	Vector name	Marker
JME4390	GGA_ LEU2ex_CrisprCas9-yl_RFP	LEU2ex
JME4393	GGA_ LYS5ex_CrisprCas9-yl_RFP	LYS5ex
JME4472	GGA_URA3ex_CrisprCas9-yl_RFP	URA3ex
JME4580	GGA_HPHex_CrisprCas9-yl_RFP	Hygromycin optimized for *Y. lipolytica*
JME4599	GGA_NATex_CrisprCas9-yl_RFP	Nourseotricin optimized for *Y. lipolytica*
JME5000	GGA_EYK1ex_CrisprCas9-yl_RFP	EYK1ex

which is involved in the fatty acid degradation pathway. Its inactivation results in cells that are unable to use fatty acids as their sole carbon source (Smith et al. 2000). The *GSY1* gene encodes glycogen synthase, which is involved in glycogen synthesis. Its inactivation results in carbon storage being redirected from sugars (glycogen) to lipids (triacyl glycerol) (Bhutada et al. 2017). The *LIP2* gene encodes lipase, which is involved in external lipid degradation (Pignede et al. 2000).

gRNAs were designed for all these targets and then assembled in the CRISPR/Cas9 vector JME4390. All gene editing was performed in the strain Po1d (JMY195). Table 3 shows the editing success rate (i.e., following phenotype screening and sequence-based verification; see the Materials and Methods section). Editing success was highly variable (7–70%), and it depended on the target gene. It was also dependent on the gRNA sequence (data not shown), a finding reported in a large scale study in *Y. lipolytica* (Schwartz et al. 2019). Because it was easy to screen and had a median editing success rate, we chose to target *GSY1* using our CRISPR/Cas9 vector set. More specifically, the CRISPR/Cas9-gGSY1 vectors were used in different genetic backgrounds adapted to the six markers (Table 4). The results show that editing success was highly dependent on the marker/strain combination used and ranged from 4% (JMY195 with the HPHex marker) to 81% (JMY330 with the LEU2ex marker and JMY195 with the NATex marker). For one marker, EYK1ex, no positive clones were obtained (Table 4). For this strain/marker combination, either editing failed completely or the success rate was below 1% and thus undetectable.

In *Y. lipolytica*, homologous recombination is not very efficient. To confirm that our set of CRISPR/Cas9 vectors could help integrate DNA via homologous recombination, we compared the rate of homologous recombination with and without co-transformation by CRISPR/Cas9 vectors in a standard genetic background and in a *Y. lipolytica* strain deleted for *ku70*, which has shown improved homologous recombination (Verbeke et al. 2013). A classical disruption cassette composed of a URA3 expression cassette flanked by 1-kb homologous regions upstream and downstream from the *GSY1* gene—was used to transform Po1d and Po1d *Δku70*. Table 5 shows that using a CRISPR/Cas9 vector in tandem with the integration cassette strongly improved homologous recombination. The success rate was nearly 100% in the *Δku70* background and 83% in the Po1d background. When the cassette alone was used, the success rate was only around 15%. High success rates were obtained in a similar experiment using the same backbone plasmid (Schwartz et al. 2016). The low level of homologous recombination in *Y. lipolytica* is a drawback in standard deletion and targeted integration procedures. Moreover, the need for a long flanking region complicates cloning, but reducing the size of the homologous region strongly reduces the rate of homologous recombination. To determine if using a CRISPR/Cas9 vector in tandem with a shortened homologous region could improve the rate of correct integration, similar experiments were performed using homologous regions of different lengths (100 bp and 50 bp) located both upstream and downstream from the *GSY1* gene. When the 100-bp region was used in the absence of the CRISPR/Cas9 vector, no positive clones resulted from the deletion of the *GSY1* gene (Table 5). However, when the CRISPR/Cas9 vector was employed, the success rate was reasonable (25%); it was very high (88%) in the *Δku70* background. When the 50-bp region was used, there were no positive clones (Table 5). In short, using a CRISPR/Cas9 vector makes it possible to strongly reduce the size of the homologous recombination region (down to 100 bp) and still obtain a reasonable rate of successful insertion with or without using a *Δku70* background.

Table 3 Target gene, ID, editing success rate, and the number of clones tested using the backbone vector JME4390 in the JMY195 strain

Gene	YALI ID	Editing success rate (number of clones tested)
MFE2	YALI0E15378g	7% (30)
EYK1	YALI0F01606g	17% (30)
EYD1	YALI0F01650g	7% (28)
GSY1	YALI0F18502g	20% (48)
URA3	YALI0E26741g	20% (48)
LIP2	YALI0A20350g	70% (30)

Table 4 Editing success rate for the *GSY1* gene and the number of clones tested for the different markers and strains

Vector	Marker	Strain	Editing success rate (%)	Number of transformants tested
JME4473	URA3ex	Y2033	19	48
JME4473	URA3ex	Y195	21	48
JME4392	LEU2ex	Y195	46	96
JME4392	LEU2ex	Y330	81	64
JME4425	LYS5ex	Y5211	56	96
JME4600	NATex	Y195	81	32
JME4600	NATex	Y330	37	80
JME4600	NATex	WT5	56	48
JME4600	NATex	W29	57.5	40
JME4759	HPHex	W29	45	13
JME4759	HPHex	Y195	4	48
JME5019	EYK1ex	Y7123	0	48

Table 5 Rate at which the *GSY1* gene was successfully edited and the number of clones tested in different genetic contexts employing homologous recombination regions of different lengths

	1-kb homologous flanking region		100-bp homologous flanking region		50-bp homologous flanking region	
	Number of positive clones	Editing success rate (%)	Number of positive clones	Editing success rate (%)	Number of positive clones	Editing success rate (%)
CRISPRgsy1	10/48[*]	20[*]	N/A	N/A	N/A	N/A
HR (PUTgsy1)	7/48	15	0/48	0	0/48	0
CRISPRgsy1 + HR (PUTgsy1)	40/48	83	12/48	25	0/48	0
CRISPRgsy1 + HR (PUTgsy1) + delta KU	48/48	100	42/48	88	0/48	0

[*]In this case no disruption cassette was used, only the efficiency of the CRISPR-Cas9 system was evaluated as a control

N/A Not applicable

One of the advantages of having access to a large set of markers is that it becomes faster to carry out multiple deletions in a single strain. We therefore wished to verify that we could edit multiple genome locations by transforming yeast using two CRISPR/ Cas9 plasmids in tandem. In the strain JMY5217 (Po1d Lys⁻Leu⁻), we simultaneously disrupted *gsy1* and *ura3* using the JME 4392 (CrCas9-Leu2-gGSY) JME 4453 (CrCas9-Lys5-gURA), respectively. For the 48 colonies tested, the success rate was 50% for

ura3 and 85% for *gsy1*. All the *ura3* knockouts were also *gsy1* knockouts. When the markers were inversed, by using JME4425 (CrCas9-Lys5-gGSY) and JME 4452 (CrCas9-Leu2-gURA), our success rate was nil for *ura3* and 62% for *gsy1*. These results show that, in *Y. lipolytica*, it is possible to carry out simultaneous transformation employing multiple CRISPR/Cas9 vectors that have the same backbone but that express different markers. However, some gRNA-marker combinations appear less efficient than others as already observed (Table 4). In particular, Leu2-gURA was not successful in our hand in the JMY5217 background while it is in the JMY195 background (Table 3).

Genome editing in wild-type strains using CRISPR/Cas9 vectors

Most of the genetic engineering that takes place in *Y. lipolytica* involves a small subset of laboratory strains that have been specifically developed for this purpose by introducing auxotrophies. Because standard engineering methods therefore rely on auxotrophies, limited use has been made of the diverse characteristics displayed by different *Y. lipolytica* wild-type strains. However, as mentioned in the introduction, wild-type strains can outperform laboratory strains. They are therefore a better chassis for carrying out further modifications. One of our objectives in developing this set of CRISPR/Cas9 vectors with dominant markers was to be able to perform genome editing in wild-type *Y. lipolytica* strains. We used our CRISPR/Cas9-hph vector, into which *gsy1* gRNA had been introduced, to transform a collection of wild-type strains as a proof of principle. This plasmid was used to transform nine different wild-type strains representing a broad range of origins (see Supplementary Table S1). The rate at which we successfully disrupted *gsy1* was determined using Lugol staining, as shown in Fig. 2 (A and B); the editing success rates are indicated in Table 6. Positive clones were sequenced to confirm editing success.

The editing success rate was determined based on the staining patterns of the transformants (dark vs. clear in response to Lugol staining). As variability in staining patterns was observed among strains (Fig. 2), we sequenced between one and three clear clones for each strain to validate the genome editing process. In all cases, we observed that genome editing occurred

upstream from the PAM sequence, at the gRNA target sequence. In general, our CRISPR/Cas9 vectors with dominant markers were well suited to genome editing in wild-type strains. All the wild-type *GSY1* genes were sequenced, and their sequences were compared to the CLIB122 sequence that was used to design the gRNA. No differences were seen in the gRNA target region, excluding potential mismatch bias. We observed dramatic differences in editing success: it ranged from 100% for the strains IMUFRJ 50682 and CBS 6125 to 0% for strains CLIB 791 and DBVPG 4400 (Table 6). This observation implies that it might be easier to genetically engineer some wild-type strains than others, but wild-type strains could still outperform some of the more widely used laboratory strains. We were able to easily transform all the wild-type strains with our vector even if two of them failed to show any signs of editing. All the strains tested are sensitive to hygromycin (growth inhibition between 60 and 80 μg/mL) and no major differences in sensitivity were observed that could explained differences in transformability and/or editing.

Discussion

Here, we describe how we used the Golden Gate method to assemble a set of CRISPR/Cas9 vectors, each containing a different marker, that can be employed to perform genome editing in *Y. lipolytica*. The vectors are now available via Addgene (129,656–129,661). In all these vectors, any gRNA sequence can be easily and rapidly introduced using Golden Gate assembly.

We built our system by taking advantage of the large set of bricks available for *Y. lipolytica* and a recently published Golden Gate protocol (Larroude et al. 2019). In our system, the Cas9 endonuclease could be placed under the control of different promoters that vary in strength and inducibility, which could be useful to temporally control expression. We decided to employ the 8UASpTEF promoter because our comparison of the results obtained with pTEF and 8UASpTEF found no significant differences in editing efficiency (data not shown). Borsenberger et al. showed that promoter strength was not critical and that pTEF provides adequate results compared to 8UASpTEF when it comes to the expression of the Cas9 endonuclease in *Y. lipolytica* (Borsenberger et al.
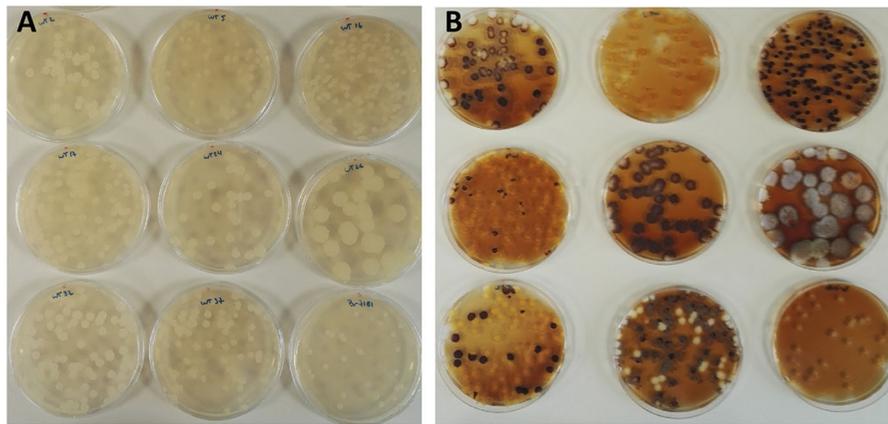
Fig. 2 CRISPR/Cas9-mediated disruption of the *GSY1* gene in wild-type strains. **a** Clones on plates before Lugol staining. **b** Clones on plates after Lugol staining. List of strains tested (from top left to bottom right): CBS 2074, CBS 6125, CLIB 791, CLIB 879, DBVPG 4400, DBVPG 5851, NCYC 3271, PYCC 4743 and IMUFRJ 50682

Table 6 Number of clones with and without the clear colony phenotype, number of positive clones sequenced, and editing success rate

Strain	Colonies with clear phenotype/total clones	Positive clones/total clones sequenced	Editing success rate (%)
IMUFRJ 50682	48/48	3/3	100
CBS 2074	9/48	1/1	18.75
CBS 6125	48/48	3/3	100
CLIB 879	90/105	3/3	85
DBVPG 5851	1/26	1/1	3.8
NCYC 3271	42/48	3/3	87.5
PYCC 4743	18/48	1/1	37.5
CLIB 791	0/153	0	0
DBVPG 4400	0/46	0	0

2018). The only constraint is that RNA structure must be compatible. Egermeier et al. (Egermeier et al. 2019) recently published a Golden Gate protocol for building *Y. lipolytica* CRISPR/Cas9 vectors, but its applicability is more limited than that of our protocol because it does not include a range of markers or a rapid-screening method (i.e., we used RFP as a negative reporter of assembly success). They were able to knockout *LEU2* in a wild-type strain and showed that their system was functional (editing success rate: 6–25%). They used a different CRISPR/Cas9 system that is based on the HH ribozyme and a humanized Cas9 that is not codon optimized for *Y. lipolytica* (Gao et al. 2016). Systems utilizing HH ribozyme gRNA processing have been shown to be less efficient than systems utilizing the PolIII promoter system (Schwartz et al. 2016); we used the latter system here, and it was coupled with a *Y. lipolytica* optimized Cas9. Accordingly, in *Y. lipolytica*, codon-optimized Cas9 is expressed at higher levels than the humanized Cas9 when the polIII system is employed (Borsenberger et al. 2018).

We also used CRISPR/Cas9-mediated cutting to introduce an expression cassette at a specific locus. The editing success rate was much higher with CRISPR/Cas9 than without CRISPR/Cas9, and we

were able to drastically reduce the length of the homologous recombination region without having to use a Δ*ku* background. This approach allowed us to avoid the pitfalls associated with the Δ*ku* background. It can also simplify and speed up metabolic engineering because integrating a large pathway at a specific locus can sometimes be a difficult task (Schwartz et al. 2017b), and using a marker can drastically improve efficiency. Marker-free integration, in which Cas9 cuts are repaired via homologous recombination, can work in *Y. lipolytica*, but the editing success rate is not consistent outside of the Δ*ku* background (Gao et al. 2016; Schwartz et al. 2017b; Holkenbrink et al. 2018).

All the vectors that we built were tested using different targets, and genome editing was successful in all cases but one (the EYK1ex marker). We took advantage of antibiotic markers to exploit the natural diversity of *Y. lipolytica* and validated our CRISPR/Cas9 vectors in a large number of wild-type strains. The editing success rate for *gsy1* disruption differed dramatically among strains even though they harboured an identical *gsy1* sequence. The colony morphology are diverse (Fig. 2) and may reflect physiology differences between strains. This can affect the rate of transformants if, for example, cell wall structures are different. Transformation procedure is the one setup and optimized for the laboratory strains W29 or CLIB122. Physiology differences may require different optimization for the other wild-type strains, which could improve transformation efficiency and ultimately editing efficiency. The phenotype screening also highlighted differences in staining patterns that probably reflect differences in glycogen accumulation. In *Y. lipolytica*, glycogen storage is detrimental to lipid accumulation (Bhutada et al. 2017), so these results indicate that *gsy1* disruption (dark-stained strains) may have a great impact on lipid production. Another, less likely, hypothesis is that the penetration of Lugol's iodine varied among strains.

To our knowledge, our set of CRISPR/Cas9 vectors is the most extensive to date for carrying out genome editing in *Y. lipolytica*.

References

Barth G, Gaillardin C (1996) The dimorphic fungus Yarrowia lipolytica. In: Wolf K (ed) Genetics, biochemistry and molecular biology of non conventional yeasts. Springer Verlag, Heidelberg, pp 313–388

Bhutada G, Kavšcek M, Ledesma Amaro R, Thomas S, Rechberger GN, Nicaud J-M, Natter K (2017) Sugar versus fat: elimination of glycogen storage improves lipid accumulation in *Yarrowia lipolytica*. FEMS Yeast Res 17(3):fox020. https://doi.org/10.1093/femsyr/fox020

Borsenberger V, Onésime D, Lestrade D, Rigouin C, Neuvéglise C, Daboussi F, Bordes F (2018) Multiple parameters drive the efficiency of CRISPR/Cas9-induced gene modifications in *Yarrowia lipolytica*. J Mol Biol 430:4293–4306. https://doi.org/10.1016/j.jmb.2018.08.024

Carly F, Gamboa Melendez H, Vandermies M, Damblon C, Nicaud J-M, Fickers P (2017) Identification and characterization of EYK1, a key gene for erythritol catabolism in *Yarrowia lipolytica*. Appl Microbiol Biotechnol 101:6573–6596. https://doi.org/10.1007/s00253-017-8412-4

Carly F, Steels S, Telek S, Vandermies M, Nicaud J-M, Fickers P (2018) Identification and characterization of EYD1, encoding an erythritol dehydrogenase in *Yarrowia lipolytica* and its application to bioconvert erythritol into erythrulose. Bioresour Technol 247:963–969. https://doi.org/10.1016/j.biortech.2017.09.168

Carsanba E, Papanikolaou S, Fickers P, Erten H (2019) Screening various *Yarrowia lipolytica* strains for citric acid production. Yeast 36:319–327. https://doi.org/10.1002/yea.3389

Celinska E, Ledesma-Amaro R, Larroude M, Rossignol T, Pauthenier C, Nicaud JM (2017) Golden gate assembly system dedicated to complex pathway manipulation in *Yarrowia lipolytica*. Microb Biotechnol 10:450–455. https://doi.org/10.1111/1751-7915.12605

DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM (2013) Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res 41:4336–4343. https://doi.org/10.1093/nar/gkt135

Egermeier M, Russmayer H, Sauer M, Marx H (2017) Metabolic flexibility of *Yarrowia lipolytica* growing on glycerol.

Front Microbiol 8:49. https://doi.org/10.3389/fmicb.2017.00049

Egermeier M, Sauer M, Marx H (2019) Golden gate based metabolic engineering strategy for wild-type strains of *Yarrowia lipolytica*. FEMS Microbiol Lett 366(4):fnz022. https://doi.org/10.1093/femsle/fnz022

Fickers P, Le Dall MT, Gaillardin C, Thonart P, Nicaud JM (2003) New disruption cassettes for rapid gene disruption and marker rescue in the yeast *Yarrowia lipolytica*. J Microbiol Methods 55:727–737

Fournier P, Guyaneux L, Chasles M, Gaillardin C (1991) Scarcity of ars sequences isolated in a morphogenesis mutant of the yeast *Yarrowia lipolytica*. Yeast 7:25–36. https://doi.org/10.1002/yea.320070104

Gao S et al (2016) Multiplex gene editing of the *Yarrowia lipolytica* genome using the CRISPR-Cas9 system. J Ind Microbiol Biotechnol 43:1085–1093. https://doi.org/10.1007/s10295-016-1789-8

Gao D, Smith S, Spagnuolo M, Rodriguez G, Blenner M (2018) Dual CRISPR-Cas9 cleavage mediated gene excision and targeted integration in *Yarrowia lipolytica*. Biotechnol J 13:e1700590. https://doi.org/10.1002/biot.201700590

Holkenbrink C, Dam MI, Kildegaard KR, Beder J, Dahlin J, Belda DD, Borodina I (2018) EasyCloneYALI: CRISPR/Cas9-based synthetic toolbox for engineering of the yeast *Yarrowia lipolytica*. Biotechnol J 13:e1700543. https://doi.org/10.1002/biot.201700543

Kretzschmar A, Otto C, Holz M, Werner S, Hubner L, Barth G (2013) Increased homologous integration frequency in *Yarrowia lipolytica* strains defective in non-homologous end-joining. Curr Genet 59:63–72. https://doi.org/10.1007/s00294-013-0389-7

Larroude M, Rossignol T, Nicaud JM, Ledesma-Amaro R (2018) Synthetic biology tools for engineering *Yarrowia lipolytica*. Biotechnol Adv 36:2150–2164. https://doi.org/10.1016/j.biotechadv.2018.10.004

Larroude M, Park Y-K, Soudier P, Kubiak M, Nicaud J-M, Rossignol T (2019) A modular Golden Gate toolkit for *Yarrowia lipolytica* synthetic biology. Microb Biotechnol 12:1249–1259. https://doi.org/10.1111/1751-7915.13427

Ledesma-Amaro R, Nicaud J-M (2016) *Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids. Prog Lipid Res 61:40–50. https://doi.org/10.1016/j.plipres.2015.12.001

Madzak C (2015) *Yarrowia lipolytica*: recent achievements in heterologous protein expression and pathway engineering. Appl Microbiol Biotechnol 99:4559–4577. https://doi.org/10.1007/s00253-015-6624-z

Morse NJ, Wagner JM, Reed KB, Gopal MR, Lauffer LH, Alper HS (2018) T7 polymerase expression of guide RNAs in vivo allows exportable CRISPR-Cas9 editing in multiple yeast hosts. ACS Synthetic Biology 7:1075–1084. https://doi.org/10.1021/acssynbio.7b00461

Nicaud JM (2012) Yarrowia lipolytica. Yeast 29:409–418. https://doi.org/10.1002/yea.2921

Nicaud JM, Fabre E, Gaillardin C (1989) Expression of invertase activity in *Yarrowia lipolytica* and its use as a selective marker. Curr Genet 16:253–260

Park Y-K et al (2019) Engineering the architecture of erythritol-inducible promoters for regulated and enhanced gene expression in *Yarrowia lipolytica*. FEMS Yeast Res 19(1):foy105. https://doi.org/10.1093/femsyr/foy105

Pignede G, Wang H-J, Fudalej F, Seman M, Gaillardin C, Nicaud J-M (2000) Autocloning and amplification of LIP2 in *Yarrowia lipolytica*. Appl Environ Microbiol 66:3283–3289. https://doi.org/10.1128/aem.66.8.3283-3289.2000

Quarterman J, Slininger PJ, Kurtzman CP, Thompson SR, Dien BS (2017) A survey of yeast from the *Yarrowia* clade for lipid production in dilute acid pretreated lignocellulosic biomass hydrolysate. Appl Microbiol Biotechnol 101:3319–3334. https://doi.org/10.1007/s00253-016-8062-y

Schwartz CM, Hussain MS, Blenner M, Wheeldon I (2016) Synthetic RNA polymerase III promoters facilitate high-efficiency CRISPR–Cas9-mediated genome editing in *Yarrowia lipolytica*. ACS Synthetic Biology 5:356–359. https://doi.org/10.1021/acssynbio.5b00162

Schwartz C, Frogue K, Ramesh A, Misa J, Wheeldon I (2017a) CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in *Yarrowia lipolytica*. Biotechnol Bioeng 114:2896–2906. https://doi.org/10.1002/bit.26404

Schwartz C, Shabbir-Hussain M, Frogue K, Blenner M, Wheeldon I (2017b) Standardized markerless gene integration for pathway engineering in *Yarrowia lipolytica*. ACS Synthetic Biology 6:402–409. https://doi.org/10.1021/acssynbio.6b00285

Schwartz C, Curtis N, Lobs AK, Wheeldon I (2018) Multiplexed CRISPR activation of cryptic sugar metabolism enables *Yarrowia lipolytica* growth on cellobiose. Biotechnol J 13:e1700584. https://doi.org/10.1002/biot.201700584

Schwartz C et al (2019) Validating genome-wide CRISPR-Cas9 function improves screening in the oleaginous yeast *Yarrowia lipolytica*. Metab Eng 55:102–110. https://doi.org/10.1016/j.ymben.2019.06.007

Shaw AJ et al (2016) Metabolic engineering of microbial competitive advantage for industrial fermentation processes. Science 353:583–586. https://doi.org/10.1126/science.aaf6159

Smith JJ, Brown TW, Eitzen GA, Rachubinski RA (2000) Regulation of peroxisome size and number by fatty acid β-oxidation in the yeast *Yarrowia lipolytica*. J Biol Chem 275:20168–20178. https://doi.org/10.1074/jbc.M909285199

Trassaert M, Vandermies M, Carly F, Denies O, Thomas S, Fickers P, Nicaud J-M (2017) New inducible promoter for gene expression and synthetic biology in *Yarrowia lipolytica*. Microb Cell Fact 16:141. https://doi.org/10.1186/s12934-017-0755-0

Tsakraklides V et al (2018) High-oleate yeast oil without polyunsaturated fatty acids. Biotechnol Biofuels 11:131. https://doi.org/10.1186/s13068-018-1131-y

Vandermies M, Denies O, Nicaud J-M, Fickers P (2017) EYK1 encoding erythrulose kinase as a catabolic selectable marker for genome editing in the non-conventional yeast *Yarrowia lipolytica*. J Microbiol Methods 139:161–164. https://doi.org/10.1016/j.mimet.2017.05.012

Verbeke J, Beopoulos A, Nicaud J-M (2013) Efficient homologous recombination with short length flanking fragments in Ku70 deficient *Yarrowia lipolytica* strains. Biotechnol Lett 35:571–576. https://doi.org/10.1007/s10529-012-1107-0

Wagner JM, Williams EV, Alper HS (2018) Developing a piggyBac Transposon system and compatible selection markers for insertional mutagenesis and genome engineering in *Yarrowia lipolytica*. Biotechnol J 13:e1800022. https://doi.org/10.1002/biot.201800022

In addition to the published results, other approaches were tested with the objective of expanding the utility of CRIPSR-Cas system and the toolbox of *Y. lipolytica*. In order to avoid selection marker, which are limited and need to be recycled (marker recovery by Cre/Lox system), marker free insertion cassettes were constructed and tested in the presence of CRISPR-Cas9. As the efficiency of HR was increased when using CRISPR-Cas9, it was expected that the system also will help the marker-free insertion. For this purpose, several strategies were set up, using *GSY1* as target locus of insertion as it proved to be efficiently disrupted by our CRISPR-Cas9 system and it is easy to screen (Larroude et al., 2020). Marker-free cassettes (composed of: genome homologous sequence-Promoter-Gene-Terminator-genome homologous sequence) were: (i) constructed with 1000 bp of homologous sequences at each side of the Cas9 cutting site (in the ORF), and also with homologous regions upstream and downstream the ORF ("far" from the cutting site); (ii) constructed on several sizes: small cassettes bearing only one gene (the fluorescent RedStarII protein or *URA3*) and longer cassettes with three genes that were very easy to screen as carotenoids production or xylose utilisation pathways cassettes; (iii) transformed as free digested cassettes and also as part of a replicative vector; transformed together with CRISPR-Cas9-Vector or in a strain already containing the Cas9-vector (two-step transformation). However, after all these tests, none of the tested conditions resulted in successful insertion at the target locus. In other published works, the markerless insertion in *Y. lipolytica* was only achieved on Δ*ku*70 strains where the NHEJ is impaired (S. Gao et al., 2016; Holkenbrink et al., 2018). Schwartz and colleagues screened 17 loci for heterologous gene integration by HR mediated by CRISPR-Cas9. In only 5 of them markerless cassettes were integrated with 50% efficiency. When using a Δ*ku*70 background, the integration at the other 12 non-efficient sites increased from 0 to 7-28% while no effect was observed on the 5 efficient sites (Schwartz et al., 2017b). Therefore, the marker free genome integration is still a limitation on *Y. lipolytica*.

Another approach tested in an attempt to further enlarge the toolbox for genome editing dedicated to *Y. lipolytica*, was the CRISPR-Cpf1 system. As mentioned in section 1.3, CIRPSR-Cas9 and CRISPR-Cpf1 differ on the endonuclease used, thus distinct features can be exploited. While Cas9 enzyme recognizes purine rich PAMs (NGG) and cleaves target DNA upstream of PAMs to generate blunt-ends DBSs, Cpf1 recognises T-rich PAMs (TTTN) and cuts DNA downstream of it, generating ~5 nt sticky

ends. The diversity on recognized PAM regions allows to target different regions of the genome. Additionally, Cas9 cuts close to the PAMs site, 3~5 nt, and thus most of the time the cleavage lead to a modification of the PAM inhibiting further cutting cycles. On the contrary, PAM site is likely being retained on Cpf1 system as this endonuclease cuts ~20 nt away and downstream from the PAM, allowing another editing event and leading to improved editing efficiency. The processing of crRNA is also different between both systems, Cpf1 only uses 20 nt DRS (directed repeat sequence) instead of the 80 nt tracrRNA from Cas9 system, which may have an influence on folding and delivery efficiency of the gRNA. (Zetsche et al., 2015) (Figure 16).

Therefore, a strategy for the use of CRISPR-Cpf1 system was design. The approach used for the construction of this system is based on a versatile and modular vector, assembled by GG, using the same design than for Cas9 system (Figure 17). The vector holds the Cpf1 enzyme and its promoter, the crRNA, selection marker and *Y. lipolytica* ARS for maintenance and propagation into the yeast and the bacterial elements for propagation in *E. coli*. As for Cas9 system, promoters, markers and crRNA are easily changed by BsaI and BsmBI enzymes. Likewise, the platform vectors contain a reporter gene (RFP) to ease the selection screening after crRNA assembly. The sequence coding the FnCpf1 enzyme (from *Francisella novicida*) was codon optimized for *Y. lipolytica* usage and fused to SV40 as NLS, needed to address the enzyme into the nucleus, and CYCt as terminator.

The crRNA expression cassette was designed based on the work of Swiat and colleagues, and comprises RNA polymerase III SNR52 promoter, the 25-nt target sequence flanked by 19 nt direct repeats (DRS) and a terminator (Swiat et al., 2017). The 25 nt spacer were choosen from *Y. lipolytica* genome by CHASSY partners at TUDelft (Jean-Marc Daran group, Netherlands), where they are developing a software dedicated to the design of these sequences, based on criteria also described by Swiat and colleagues. The SNR52 promoter used in our system was the same than in Cas9 system as well as the Poly-T terminator sequence.

The Cpf1 system was developed, the platform vectors with different markers and vectors containing crRNA targeting *GSY* and *URA3* were constructed. The system was tested, first results showed around 12% of disruption of *GSY* using a Cfp1-hph vector and around 2% when using *LEU2* as marker. On the other hand, no disruption could be

detected on the *URA3* gene, but tests are still on-going and the system is still under development.

In the meantime, the Cpf1 system was very recently published by Yang and co-workers, achieving editing efficiencies of 40-90% depending on the target site and were able to target two and three genes at the same time. They use the same PolIII system and DRS sequence that the one in my work, however they only use one DRS instead of two (Yang et al., 2019).

3.1.3 Discussion and conclusion

To summarize, an efficient, versatile and modular method for DNA assembly, allowing the construction of multigene and combinatorial expression cassettes for *Y. lipolytica*, was developed based on GG strategy. A scaffold of predesigned 4 nt overhangs covering three TUs (each of these expression cassettes consisting of a promoter, a gene and a terminator), selection marker gene and genomic integration targeting sequences, constituting altogether thirteen elements, was created. A broad library of interchangeable building blocks has been constructed, characterized and is available to the scientific community together with a detailed protocol to efficiently design new parts and perform specific assemblies. In addition, the combinatorial approach was efficiently used to screen the best promoter-gene pairs. Finally, the importance of this tool on the construction of metabolically modified strains was evidenced by the fast construction of strains with complex heterologous metabolisms and by its rapid handling within the team for different projects.

Even though the design and construction of the building blocks can be time consuming, the assembly of DNA parts, done in one step reaction, is very fast and highly efficient once the library is constructed, thus, greatly reducing the time for the construction of strains with complex metabolic pathways. In addition, the available library of building blocks can be expanded limitless with sequences from endogenous or heterologous origins, thus, increasing the combinations of possible assemblies to be done and the number of strains to be constructed and tested. This is of great importance for the identification of best producer strains when constructing cell factories, but also for deciphering still not-known, or not well-known, traits of this yeast. In addition, the system is well suited for combinatorial approaches, permitting highly efficient shuffling of parts, which is particularly interesting for the construction of libraries in a fast and easy

mode. This combinatorial approach, followed by the screening of desired phenotypes, is a very effective way to identify strains with the best desired phenotype in a high throughput manner. Furthermore, the multi-gene assembly capacity of the system has the advantage of reducing the rounds of transformations needed to reach a desired construction, which increases the viability and/or transformability of recombinant strains. Moreover, the fact that the cassettes can be designed to be inserted at a specific site in the genome is of great help when constructing metabolic engineered strain that need pathways to be overexpressed at the same time that deleting a competing pathways. Additionally, the site-directed modifications allow the construction of strains better characterised and defined, which is required for industrial yeasts. Finally, as the design is standardized, collaboration among different groups can be easily conducted by sharing the compatible building blocks.

Regarding CRISPR-Cas9 system, a strategy for the construction of versatile and modular vectors permitting the reutilization of GG building block was designed, and a detailed protocol was setup. A set of platform vectors with different markers, ready to add any gRNA and to be used in different background strains, was constructed and is publicly available. In addition to genes disruption, the system proved to increase the insertion of expression cassettes at a desired locus, even with homologous sequences as small as 100 bp. However, markerless insertion is still a deficient approach that need further work in order to make it properly functional. With the developed GG vector construction system, markers are very easy to change and Cas9 can be placed under the control of different promoters, including inducible ones, which could be useful for temporal control expression of this endonuclease. The same modular strategy was setup for a CRISPR-Cpf1 system. First results showed that the system is functional, however, efficiency is still low. This low disruption rate can be due to the target site, gRNA secondary structures, and also related to strain and markers used as already evidenced for Cas9. Thus, works on this system are still ongoing in order to improve it.

The CRISPR system has revolutionized genome editing of biological systems due to its easy to use technology, high specificity and capacity to target almost any sequence on the genome. Even though, efforts are still needed in order to better understand the efficiency of gRNA, that showed to be dependent on the strain and on the target sequence but whose predictability is not optimal yet. In addition to transcriptional activation and repression which already proved to work in several organisms including *Y.*

lipolytica, the functionality of the system can be enlarged. For instance, modulation of epigenetics marks and modulation of genome architecture could be conducted as in principle all the elements needed , such as histone acetylases, methylases and kinases, can be recruited by Cas9 (Mali et al., 2013). Furthermore, the development of systems using other Cas enzymes can increase the scope of sequences to be targeted, and can have an impact on the efficiency, specificity and ease to use.

The development of robust, efficient and predictable biology tools is essential to enlarge the scope of metabolic engineering possibilities in any organism. In *Y. lipolytica*, the fast and combinatorial assembly of complex synthetic pathway together with the availability of an efficient and site-specific genomic editing technique greatly enriches the spectrum of possible studies that can be conducted in this industrially relevant microorganism. However, further improvement and expansion of the genetic toolbox is possible. In this sense, the development of stable episomal plasmids, which are useful to rapid and temporary express genes, for instance for biosensor development, is still a debt in *Y. lipolytica* toolbox. Also, efforts are still needed to improve CRISPR technology, in particular on deciphering the efficiency of gRNA, on enhancing the efficiency of multiplex technology and on markerless insertion cassettes. The development of new selection markers, allowing to extend the round of genetic modification, as well as further identification and characterisation of terminators for the construction of expression cassettes, are part of the list of targets to be improved in *Y. lipolytica* toolbox. These remarks notwithstanding, the last emerging tools have demonstrated the potential of synthetic biology to revolutionize metabolic engineering efforts and to expand the possibilities of this yeast as efficient industrial cell factory. These new developed techniques were used for the construction of strains developed on the second part of this work (Section 3.2).

3.2   **Metabolic engineering of the aromatic amino acid pathway**

Metabolic engineering of microorganisms enables the development of high-performant strains for the efficient production of a wide variety of value-added compounds from renewable non-food biomass. However, the way to go to industrial microbial production is not straightforward. Microbial factories often require the optimization of long metabolic pathways in order to increase the productivity. An efficient way to accelerate the bioproduction of molecules is by constructing chassis strains optimized for a high supply of precursors that are common building blocks for the synthesis of a wide range of molecules. Thus, the biosynthetic pathway producing the molecule of interest can be plug into the already optimized strain and reduce the development time.

As it was previously introduced, aromatic chemicals have a broad spectrum of applications, thus, used in large amounts in various industries. Engineering microorganisms as cell factories for the production of aromatic compounds is an alternative to replace their unsustainable petroleum derived synthesis (Niu et al., 2002; Suástegui and Shao, 2016). In microorganisms, AAAs are the primary substrates for the biosynthesis of aromatic compounds. For this reason, the construction of microbial cell factory with a high available amount of AAA that could serve as chassis is of major interest.

Y. lipolytica is well studied for the production of fuels and chemicals derived from fatty acids, but relatively untapped regarding the AAA pathway and the biosynthesis of its derivatives compounds. However, the high flux through the PPP, the high acetyl-CoA and malonyl-CoA pools and the hydrophobic lipid bodies make *Y. lipolytica* a promising host for the biosynthesis of various aromatic derived compounds. For instance, the biosynthesis of flavonoids and stilbenes need three molecules of malonyl-CoA as well as P450 enzymes whose activity are favoured by the hydrophobic environment (Jiang and Morgan, 2004; Lv et al., 2019a). In fact, *Y. lipolytica* was previously used to express heterologous P450 enzymes (a human cytochrome P450 isoform), proving high expression and activity of the produced enzyme (Nthangeni et al., 2004). Moreover, this yeast lacks Crabtree effects, avoiding the loss of carbon source in ethanol.

In this chapter, I describe the setup of methods that were used for the evaluation of the AAA pathway optimization, the construction of *Y. lipolytica* chassis strains in which

the pool of AAA has been increased, the evaluation of feedback regulation of the pathway, and finally, the use of chassis strains as cell factories for the production of 2PE, naringenin, resveratrol and melanin.

3.2.1 Aromatic Amino Acids chassis strain

During this work, I aimed to construct a chassis strain with an optimized supply of Phe and Tyr. To this end, methods for measuring the changes on the AAA pool were developed and tested, and a series of strains with serial modifications of the AAA pathway were constructed.

3.2.1.1 Evaluation of AAA pathway improvements

With the aim of developing screening methods for the evaluation of the strains to be constructed in this work, three different approaches were tested: a promoter-based biosensor, the biosynthesis of a coloured reporter molecule and a HPLC method.

Promoter-base biosensor.

In a general manner, promoter-based biosensors systems consist of a regulator (detector module) that is able to sense the level of an intra- or extracellular effector molecule and activate the expression of a reporter gene (effector module) that produces a measurable output. These systems can be optimized to increase their sensibility, e.g., by protein engineering of TF or increasing of activator sequences on the promoter region to have a stronger synthesis of the reporter. As previously mentioned, biosensors are mostly developed in conventional organism. To my knowledge, there are no such systems developed in *Y. lipolytica*.

It was previously observed in *Y. lipolytica* that the expression of *ARO10* (YALI0D06930) was increased in the presence of Phe (Celińska et al., 2015b). Furthermore, in *S. cerevisiae* the expression of *ARO10* is activated by exogenous aromatic amino acids through Aro80p transcription factors and it is totally repressed in absence of the amino acid (Iraqui et al., 1999). Thus, it was hypothesized that the promoter region of this gene could be used to evaluate the intracellular concentration of AAAs, and a way of measuring the flux through the AAA pathway. Looking at the other two genes involved in the Ehrlich pathway, in *Y. lipolytica, ARO8* was reported as constitutively expressed and *ARO9* was not detected in the proteome analysis of Celinska *et al.*

(Celińska et al., 2015b). In *S. cerevisiae*, *ARO8* expression is subject to the general control of amino acid biosynthesis, with Gcn4p as probably acting on the basal level transcription, but no pathway-specific regulation was observed. The expression of *Aro9* in *S. cerevisiae*, on the other hand, is induced through Aro80p, similarly to ARO10, when aromatic amino acids are present in the growth medium (Iraqui et al., 1999, 1998). The *Y. lipolytica* promoter regions of these genes were also included in the test because they are involved in the metabolism of AAAs and may induce some response on the system, at least at the transcriptional level. Additionally, they can be used as controls to compare the response of *ARO10* promoter.

To construct the biosensor system, the promoter regions of these genes were fused to a reporter red fluorescent protein (RedStarII) by GG strategy. The promoter regions were arbitrarily defined as 1000 bp upstream the ORF of the gene and were amplified from the genome of W29 strain by PCR. The obtained expression cassettes (Figure 18) were inserted in the genome of *Y. lipolytica.* Confirmed clones were cultured with different concentrations of AAAs. Basically, 3 clones of each construction were cultured on a 96-well microplate in YNB supplemented with 1, 0.5 or 0.1 mg/mL of Phe, Tyr or Trp. Growth (OD_{600}) and fluorescence (Ex: 558, Em: 586) were measured every 30 min during 48 h. The fluorescence was measured as an indicator of protein expression level.

Results showed that none of the tested constructions induced the expression of the fluorescent protein in the tested conditions.

Biosensor system

| ZetaU | Mkr | AroPromoter | RedStarII | T-Lip2 | ZetaD |
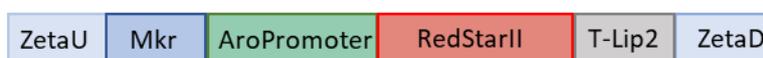
➢ *Y. lipolytica* transformation
➢ Culture in minimal media with different concentration of Trp, Tyr or Phe
➢ Measure of fluorescence of several clones

Figure 18. Aromatic biosensor construction

Promoter regions of the genes *ARO8*, *ARO9* and *ARO10* (AroPromoter) were assembled with a red fluorescent protein (RedStarII) and Tlip as terminator. The expression cassette includes a selection marker (Mkr) and Zeta sequences for insertion into the genome (ZetaU: ZetaUp, ZetaD: Zeta Down). The cassettes were transformed into *Y. lipolytica* genome. The fluorescence was measured after the addition of AAAs into the culture medium.

Looking at transcriptomic data available within the team it was noted that the expression of these three genes is at least ten times lower than *TEF* (a known highly expressed gene), and that the expression level of *ARO9* in different conditions is several times lower than the *ARO8* and *ARO10*. The lower expression of *ARO9* is in accordance with the absence of ARO9 protein in the proteomic analysis previously mentioned (Celińska et al., 2015b). This very low gene expression level may be an explanation of the lack of expression of the reporter gene using these promoters.

In Addition, an analysis of these Aro promoter sequences among the Yarrowia clade was done in order to identify conserved sequences, putative UAS that could be involved in gene expression. These sequences can be added in tandem to construct hybrid promoters with an increased activation strength, thus, raising gene expression, as previously demonstrated for UAS1 of gene *XPR2* (Madzak et al., 1999) and UAS_{eyk} of gene *EYK1* (Trassaert et al., 2017). Finally, no obvious homology was found between these sequences and no consensus sequences could be assigned.

Due to the lack of induced gene expression using these promoters and the absence of sequences that could be used to engineer them and increase their response, no further tests were done regarding this biosensor approach. Moreover, our CHASSY partner at TU-Delft (Netherlands) tested the same approach in *S. cerevisiae* and didn't succeed either.

However, in a very recent work, regulatory motifs were identified in the promotor regions of the genes *ARO3* and *TRP5* (Trebulle, 2019), and could be tested in the future as biosensors of the aromatic pathway.

PVA as an intracellular reporter of the shikimate pathway

The second approach used in this work to evidence changes on the AAA pathway is based on the biosynthesis of a reporter molecule whose precursor is part of the pathway being evaluated. The reporter molecule has to be easily detected, i.e. produce a fluorescent or coloured signal. The output signal is thus directly related with the flux of the pathway under study.

The PVA, intermediate in violacein biosynthesis (see section 1.4.5.1), is a green pigment derived from Trp. Therefore, it can be used as reporter of the AAA pathway.

For its production in *Y. lipolytica* the heterologous genes *VioA*, *VioB* and *VioE* have to be expressed.

The expression cassette, constructed by GG, was available in the laboratory (Kholany et al., 2019), and the three-gene pathway was expressed in strains differentially engineered at the AAA pathway. The differential flux through the shikimate pathway in the different constructed strains was evidenced by the production of various green intensities (Figure 19), and could be, thus, evaluated. This assay showed that the overexpression of genes upstream chorismate increase the availability of Trp for the production of PVA, resulting in cell cultures showing a darker green colour (Figure 19 B and C). On the contrary, when *ARO7* is overexpressed, the chorismate is less available for the production of Trp thus a lighter green colour is formed (Figure 19 B and D). This confirm that PVA could be as reporter of AAA flux modifications.

However, it is known that the expression level can vary depending on the insertion loci (Bordes et al., 2007; Leplat et al., 2015), introducing high variability within the different strains. Therefore, the engineered strains were constructed again by introducing first the *VioABE* expression cassette followed by the insertion of the different AAA pathway modifications. Thus, all strains tested derived from the same strain bearing the reporter system, avoiding bias due to reporter cassette integration site. The results obtained were identical to the first series of strain (data not shown), confirming that the effect was due to the modification on the AAA pathway and not because of random insertion of the PVA pathway.

The system was also tested on microplate cultures with the purpose of rapidly and efficiently measure the colour produced in a large quantity of strains. The inconvenient was the interference of cell absorbance (Abs 600nm) with the pigment colour absorbance (Abs 640nm) which did not allow a proper measure and quantification of the PVA produced.

In order to overcome this problem, a pigment extraction was tested but was not efficient enough in such small volume. Thus, leading to inaccurate measurements that did not reflect the difference that were clear by eye.

Even though the technique put into evidence the influence of the overexpressed genes in the shikimate pathway, it was not efficient as high-throughput detection method due to the difficulties to quantify it in the tested conditions. Additionally, strains

overproducing Tyr and Phe appeared as non-coloured strains because the chorismate is pulled away from the Trp, and PVA, pathway. Thus, this system is not convenient to evaluate Tyr and Phe overproducing strain. Notwithstanding, the system permitted the identification of a strain that can be used as chassis strain for the synthesis of products derived from Trp.
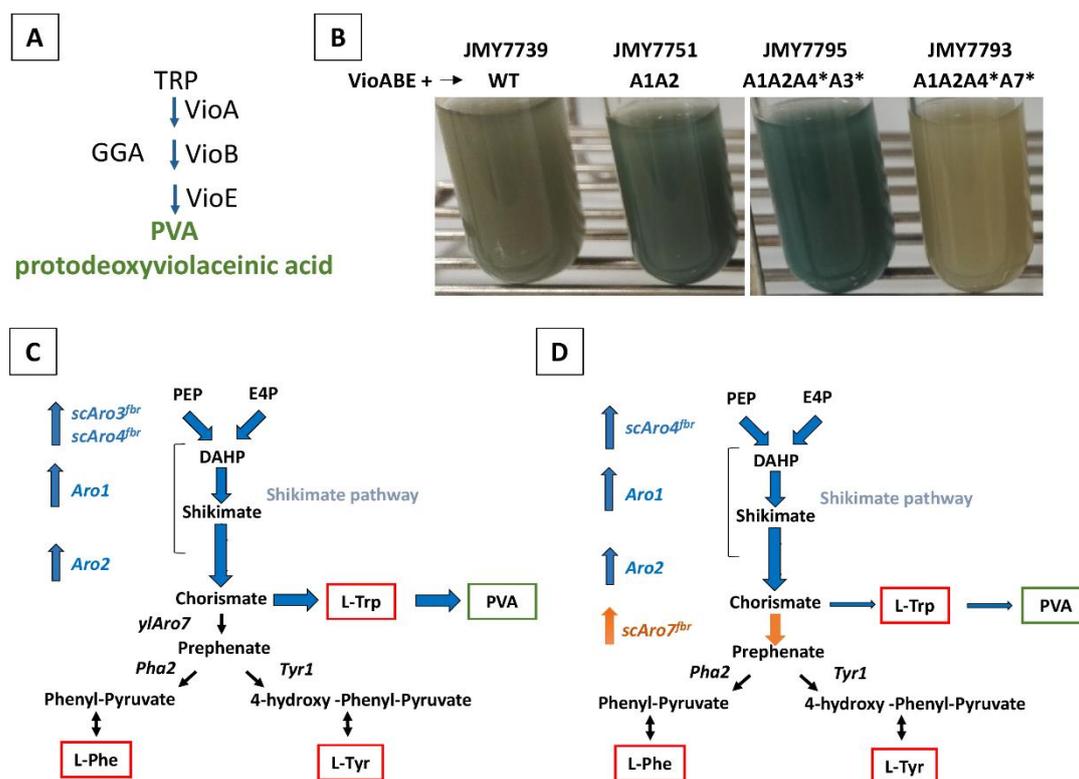
Figure 19. PVA as intracellular reporter of shikimate pathway

The PVA, intermediate in violacein biosynthesis, is a green pigment synthesized from Trp. Its use as reporter of the shikimate pathway is shown. **A-**Enzymes from the violacein pathway that were expressed in *Y. lipolytica* in order to produce PVA from Trp. A 3TU-GG Assembly (GGA) strategy was used to construct the expression cassette. **B-** Strains overexpressing vio genes as reporter and Aro genes. WT: wild-type strain expressing only *vioA-vioB-vioE*; A1A2: *Aro1-Aro2* from *Y. lipolytica*; A4*A3*: *scAro4^{K229L}-scAro3^{K222L}*; A4*A7*: *scAro4^{K229L}-scAro7^{T226I}*. **C-** Schematic representation of the shikimate pathway and how the flux is modified when genes upstream the chorismate are overexpressed. In this case the flux to Trp can be increased and so the production of PVA (A1A2A4*A3* strain) **D-** Schematic representation of the shikimate pathway and how the flux is modified when genes downstream the chorismate are overexpressed. In this case the Trp available to produce PVA decreases and less intense green cultures are obtained (A4*A7* strain). * = fbr: feedback resistant version of the enzyme (*scAro4^{K229L}, scAro3^{K222L}, scAro7^{T226}*). OE: Overexpression. sc: *S. cerevisiae*

HPLC

HPLC (high-performance liquid chromatography) is a very well-known analytical technique that allows the analysis of a sample by separation, identification and quantification of its components. Thus, using a C_{18} (non-polar stationary phase) column and a diode array detector (DAD), a methodology was set up for measuring the amount of Ehrlich metabolites, derived from the catabolism of AAA, produced by the strains. These compounds could be easily detected in the culture supernatant. The protocol of the methodology is described in section 2.4.1. This method was used to analyse the engineered strains.

However, as the production of Ehrlich metabolites was not very high on the first constructed strains, the genes coding for Aro8 and Aro10 were first overexpressed in order to increase the production of these compounds and the analytical detection signal.

Even though the method is analytically very accurate and easy to run once it is setup, it is not a high-throughput method for the rapid screening of a very large set of strains obtained, for instance, after the construction of a combinatorial library. Notwithstanding, this method resulted in a good way to evaluate the different constructed strains, allowing a quantitative measure of the pathway. Additionally, the setup HPLC method is also efficient for the detection of naringenin, resveratrol, AAAs and other intermediates of the pathway.

3.2.1.2   Construction of chassis strain

To enhance metabolic fluxes towards the AAA pathway, several well-defined conventional strategies have been employed in different engineered microorganisms: (i) increasing the activity of enzymes involved the pathway (ii) alleviating bottleneck reactions, (iii) increasing the availability of the two key shikimate pathway precursors and (iv) down-regulating competing pathways.

Based on the knowledge available from *S. cerevisiae*, and combining the strategies mentioned, *Y. lipolytica* was engineered in order to transform it into a chassis strain with an increased pool of AAA, more precisely Phe and Tyr.

First, I identified all ORFs coding for the enzymes involved in the AAA pathway in *Y. lipolytica*. To do so, the amino acid sequences of the well-characterized proteins

from the AAA pathway of *S. cerevisiae* were blasted against *Y. lipolytica* genome. The identified ORFs were in accordance with the annotation of *Y. lipolytica* genome available on GRYC and also with the information found in KEGG database (https://www.ge-nome.jp/kegg/). The list of *Y. lipolytica* genes is presented hereafter (Table 3).

Table 3. List of genes involved in the AAA pathway of Y. lipolytica.

Name	Gene ID	% identity with *S. cerevisiae* proteins	% Query cover
ARO1	YALI0F12639	51.8	98
ARO2	YALI0D17930	70.2	99
ARO3	YALI0B20020	66.6	98
ARO4	YALI0C06952	71.4	98
ARO7	YALI0E17479	55.4	100
ARO8	YALI0E20977	50.4	98
ARO9	YALI0C05258	28.8	97
ARO10	YALI0D06930	36.1	95

Afterwards, the identified genes were amplified from *Y. lipolytica* W29 genome by PCR and adapted according to our GG assembly protocol. When necessary, internal BsaI recognition sites were eliminated by site-directed mutagenesis doing a two-step PCR. When using *S. cerevisiae* genes (see later in the text), they were codon optimized according to *Y. lipolytica* codon usage bias, avoiding internal BsaI sites, and synthesized. To reduce the rounds of transformation, most of the genes were assembled by pairs using the 2-TU GG strategy. Thus, six vectors containing different gene pairs were constructed. All constructions were performed using zeta sequences for random integration in the genome, pTEF constitutive promoter for gene expression and tLip2 as terminator. Selection markers were used depending on the genotype of the strain. By combining these vectors, eighteen strains with serial modifications were constructed and evaluated in order to find the best performant chassis strain. In the following paragraphs, while describing constructions, when no indication is given genes are from *Y. lipolytica*, on the other hand, genes from *S. cerevisiae* are indicated as sc. The complete list of vectors and strains, as well as primers used and sequences of codon optimized genes, are available in Annexes (Sections 6.1.3 and 6.1.4)

As mentioned in section 3.2.1.1, a HPLC method was set up to analyse the constructed strains by the detection of Ehrlich metabolites. Celinska and co-workers previously evidenced that not all *Y. lipolytica* strains are able to produce Ehrlich metabolites at detectable levels (Celińska et al., 2013), and later identified *ARO8* and *ARO10* genes

as involved on the catabolism of AAA in *Y. lipolytica* (Celińska et al., 2015b). My experiments revealed that in the *Y. lipolytica* strain we use in our laboratory, derived from W29, *ARO8* and *ARO10* genes have to be over-expressed in order to reach a detectable level of Ehrlich metabolites by HPLC. This indirect measure of the AAA pool was chose to evaluate the constructed strains considering that no biosensor could be developed and that AAAs, from intracellular or supernatant samples, were not directly detectable by HPLC for the wild-type nor for the first strains overexpressing part of the AAA pathway.

As shown in Figure 20, the expression of only one part of the pathway, using the cassettes containing the pairs *ARO1-ARO2* or *ARO4-ARO7*, does not have any detectable effect on the Ehrlich metabolite production, but a considerable increase is observed when the whole pathway is overexpressed, by overexpressing both cassettes (*ARO1-ARO2* and *ARO4-ARO7*) simultaneously (strain 8306, Figure 20. See also Figure 22).
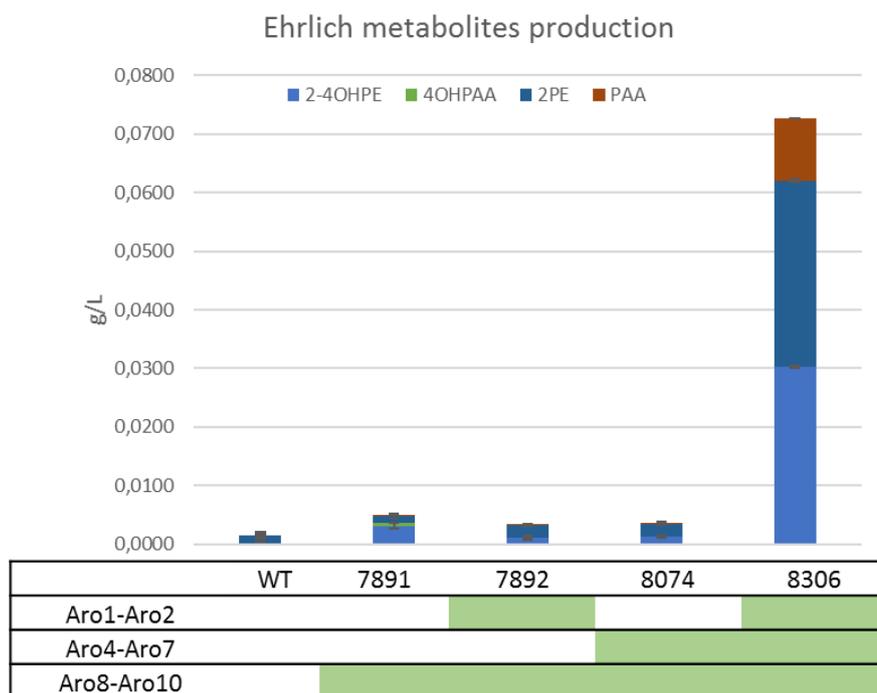
Figure 20. Evaluation of the AAA pathway of engineered strains.
Ehrlich metabolites production is used as reporter of the flux through the AAA pathway. The genotype of the constructed strains is indicated by the green-filled cases below their numbered-name. The numbers above the genotype indicate their collection number (JMYnumber). WT: wild-type strain. 2-4OHPE: 2-(4-Hydroxyphenyl)ethanol; 4OHPAA: 4-Hydroxyphenylacetic acid; 2PE: Phenylethanol; PAA: phenylacetic acid.

As previously mentioned, in *S. cerevisiae*, two reactions of the shikimate pathway are known to be feedback inhibited: (i) the first committed step of the shikimate pathway catalysed by Aro4 and Aro3, and (ii) the first branch point, catalysed by Aro7. The expression of feedback insensitive forms of these enzymes, scAro4^{K229L} and scAro7^{T226I}, was proved to increase the flux thought the shikimate pathway in this yeast (Krappmann et al., 2000; Luttik et al., 2008). Thus, with the aim of having the same effect, codon optimized versions of these enzymes were expressed in *Y. lipolytica*.

As expected, the expression of feedback insensitive *S. cerevisiae* enzymes increased the production of Ehrlich metabolites (Strain JMY7903 in Figure 21) indicating that the flux through the shikimate pathway was increased in this strain compared with the strain overexpressing only *Y. lipolytica* genes (Figure 21).

To improve the performance of the strain even further, the effect of two other enzymes was tested in the previous best strain (JMY7903): (i) The deregulated version of *S. cerevisiae* Aro3 (scARO3^{K222L}), which was suggested to be more effective than its isoenzyme Aro4 (Brückner et al., 2018; Hartmann et al., 2003), and (ii) *Y. lipolytica*
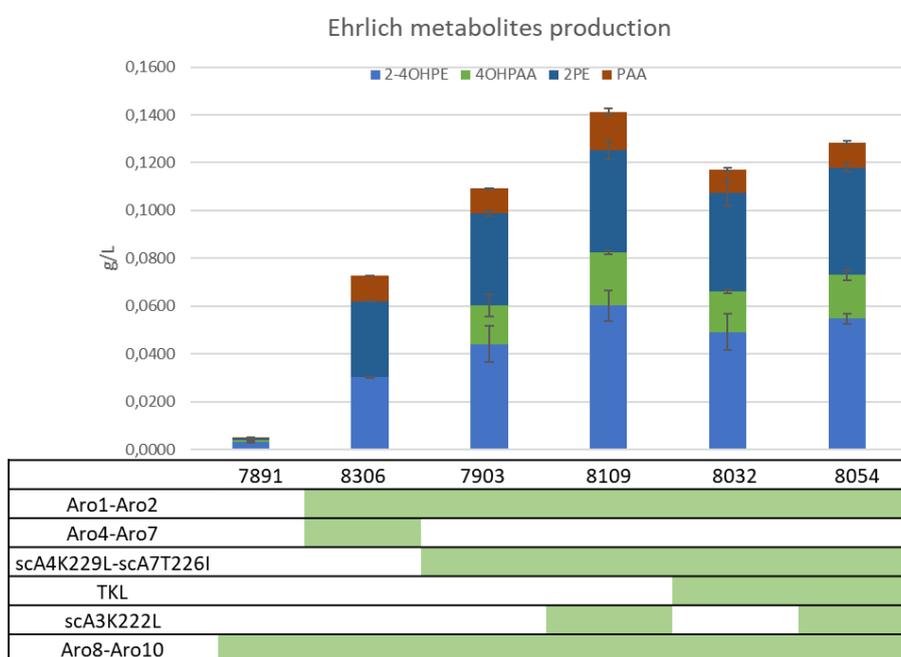
Figure 21. Evaluation of AAA the pathway on engineered strains overexpressing *S. cerevisiae* deregulated forms of Aro4 and Aro7 enzymes.

Ehrlich metabolites production is used as reporter of the flux through the AAA pathway. The genotype of the constructed strains is indicated by the green-filled cases below their numbered-name. The numbers above the genotype indicate their collection number (JMYnumber). 2-4OHPE: 2-(4-Hydroxyphenyl)ethanol; 4OHPAA: 4-Hydroxyphenylacetic acid; 2PE: Phenylethanol; PAA: phenylacetic acid

Tkl1, aiming to optimize the availability of the precursor E4P (Figure 22). The expression of scARO3^{K222L}, has a positive impact on the production of Ehrlich metabolites, meaning that the synthesis of Phe and Tyr is enhanced (Strain JMY8109 in Figure 21). On the contrary, the overexpression of *TKL1* seems to have a negative effect on this pathway flux. Even though a small increase of Ehrlich metabolites is observed after adding *TKL1* into the strain JMY7903, when scARO3^{K222L} and *TKL1* are overexpressed together the improvement is lower than when only scARO3^{K222L} is overexpressed (Figure 21).

From the series of constructed strains, JMY8109 is the one producing the highest amounts of Ehrlich metabolites on the tested conditions, implying that the synthesis of Phe and Tyr in this strain is increased compared to the wild-type strain (Figure 20 and Figure 21). This strain produces over 0.14 g/L of Ehrlich metabolites in minimal YNB medium, while the strain overexpressing only *ARO8* and *ARO10* (JMY7891) produces 0.05 g/L, demonstrating that the flux through the shikimate pathway was significantly increased.
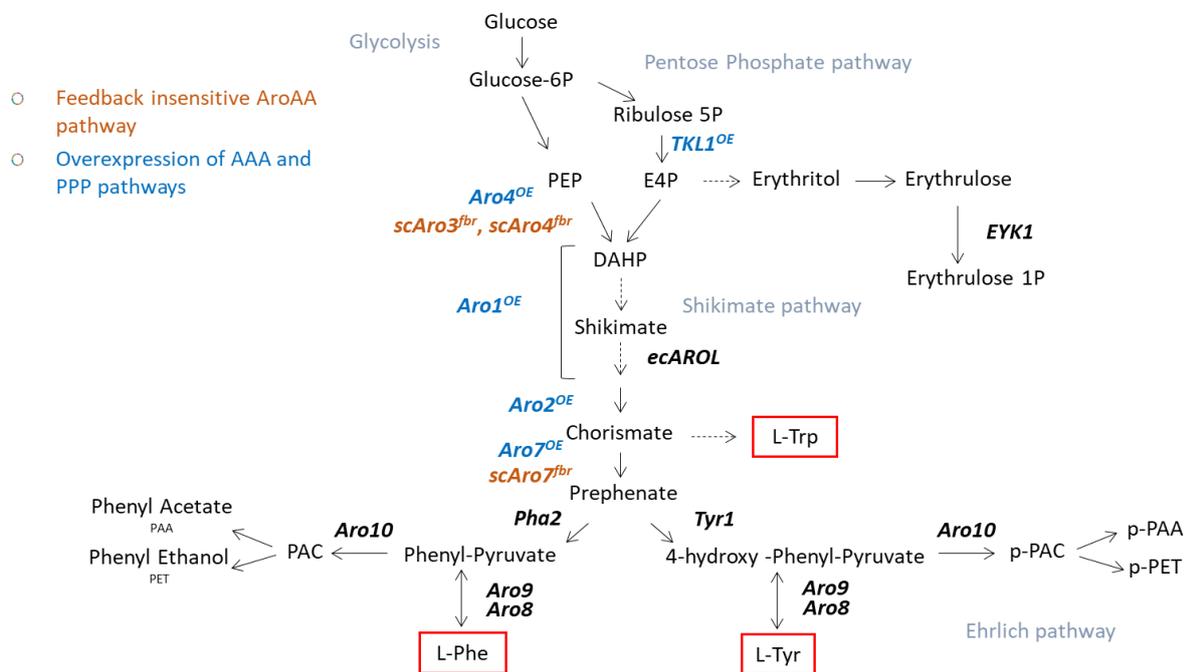
Figure 22. Pathway engineering for the construction of *Y. lipolytica* AAA chassis strain

Representation of the AAA pathway. Precursor pathways (glycolysis and Pentose phosphate pathways) and competing pathways (Erythritol and Ehrlich pathways) are also represented. Genes used, at different steps, for engineering the strains are indicated in brown and blue. Discontinued arrows indicate multi-steps reaction. OE: Overexpression of endogenous genes. fbr: Feedback insensitive mutation. sc: *S. cerevisiae.* Red squares highlight the final compounds of the pathway, the AAAs.

In order to verify the expression of the incorporated genes, and to have a better characterisation of the constructed strain, a qRT-PCR analysis was conducted on JMY8109 and the wild-type strains. Fold change calculation have been performed for *ARO*1 and *ARO*2 for which endogenous version are available. As seen in Figure 23-A, both genes are overexpressed by a factor superior to 100. However, when looking at the relative expression level ($2^{-\Delta CT}$), the *ARO*2 in the wild-type strain is 24 time lower than *ARO*1 and therefore a similar trend is observed in the chassis strain (Figure 23-B). All heterologous genes have a relative level of expression higher than *ACT1*, which reveal a strong overexpression of these genes.

Overall, the JMY8175 strain, equivalent to JMY8109 but without overexpression of *ARO8-ARO10,* can be a good chassis strain for the production of aromatic derivative compounds. This strain is the first *Y. lipolytica* AAA chassis strain. The knowledge acquired from it will feed the DBTL cycle to improve next round of construction and experiments, which will hopefully lead to the construction of more performant strains.
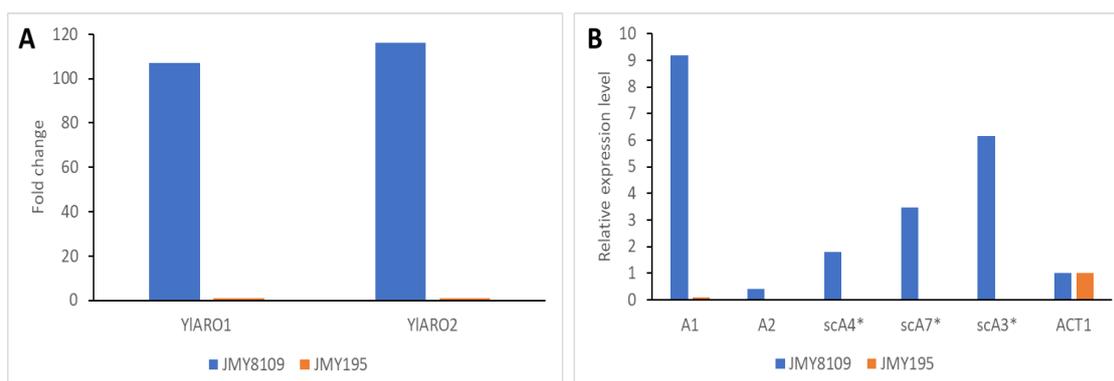
Figure 23. Gene expression in the chassis strain.

Result from a qRT-PCR analysis comparing the chassis strain and a wild-type strain are shown. A- The increase on endogenous genes expression on the chassis strain compared to the non-engineered strain. B- The expression of overexpressed genes compared to the level of expression of ACT1. A: Aro. sc: *S. cerevisiae*. Yl: *Y. lipolytica.* ACT1: gene coding for actine. *: feedback insensitive version of the enzyme.

3.2.2   Regulation of Aro4 and Aro7 in *Y. lipolytica*

It was previously mentioned that the AAA biosynthetic pathway is tightly regulated. The Aro4 and Aro7 enzymes are known to be feedback regulated by Tyr and Trp in *S. cerevisiae*, and feedback insensitive variants of these enzymes significantly increase the flux through AAA pathway (Luttik et al., 2008; Reifenrath et al., 2018).

In order to investigate this regulation in *Y. lipolytica*, the effect of Trp and Tyr on the enzymatic activity of *Y. lipolytica* Aro4 and Aro7 enzymes was analysed *in vitro*

(Figure 24). The results showed that Aro4 seems to be repressed by Tyr and Trp (Figure 24A), however, Aro7 does not seems to be affected by neither of them (Figure 24B). This suggest that the regulation of the shikimate pathway in *Y. lipolytica* might behave differently than in *S. cerevisiae*.
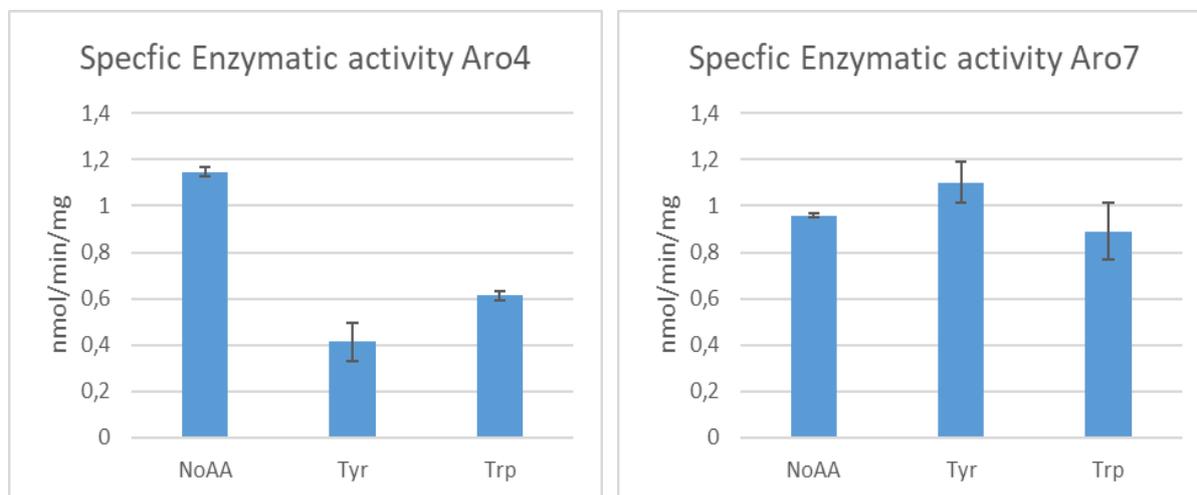
Figure 24. Specific enzymatic activity of *Y. lipolytica* Aro4 and Aro7

The enzymatic activities were measured in presence of Tyr or Trp and compared with the activity when no amino acids were present. **A**- shows the enzymatic activity of Aro4 and **B**- shows the enzymatic activity of Aro7. The test was done in duplicate. The results show the media of both test and the standard deviation between them.

To have further information about this regulation, the same mutations that were described to deregulate Aro4 and Aro7 *S. cerevisiae* enzymes were introduced in *Y. lipolytica* genes. To do so, the protein sequences of *Y. lipolytica* and *S. cerevisiae* were aligned (Figure 25). The region associated to feedback regulation in Aro4 of *S. cerevisiae* was found to be conserved in *Y. lipolytica*. Thus, the same lysine at position 221 in Aro4 of *Y. lipolytica*, corresponding to position 229 in Aro4 of *S. cerevisiae*, was converted into serine (K221L) as for *S. cerevisiae* feedback insensitive version. On the other hand, several point mutations were described to deregulate *S. cerevisiae* enzyme Aro7, being G141S and T226I the better characterized (Krappmann et al., 2000; Luttik et al., 2008; Schnappauf et al., 1998). The mutation G141S abolishes the effects of both Tyr and Trp while T226I abolishes the inhibition by Tyr but conserves the stimulation by Trp. However, when comparing protein sequences only the equivalent of G141 position was conserved in *Y. lipolytica*, thus the mutation from glycine to serine was done at this site, resulting in a Aro7 *Y. lipolytica* enzyme with the mutation G139S. Site-

directed mutations were done by PCR, using the wild-type sequence as template, and verified by sequencing.

```
                    Aro4p                                              Aro7p

Yl    8  NASSAEDVRILGYDPLLAPALLQTEVASTKNARETVSKGRKDSIDVITGKSDKLLCIVGP  67    Yl    1  MDFTKADTVLDLANIRDSLVRMEDTIVFNLIERAQFCRSEFVYKAGNS--DIPGFKGSYL  58
         N + EDVRILGYDPL +PALLQ ++ +T + ET +GR+++ID+ITGK D++L IVGP               MDFTK +TVL+L NIRD LVRMED+I+F IER+ F    VY+A +   +IP FKGS L
Sc   16  NQGAEEDVRILGYDPLASPALLQVQIPATPTSLETAKRGRREAIDIITGKDDRVLVIVGP  75    Sc    1  MDFTKPETVLNLQNIRDELVRMEDSIIFKFIERSHFATCPSVYEANHPGLEIPNFKGSLL  60

Yl   68  CSLHDPKAAMEYAQRLKELSDKLSGELVIVMRAYLEKPRTTVGWKGLINDPDMDESFNIN  127   Yl   59  DWFLQESEKVHAKLRRYAAPDEQAFFPDDLPEAILPPIDYAPILAPYSKEVSVNDEIKKI  118
         CS+HD +AA EYA RLK+LSD+L G+L I+MRAYLEKPRTTVGWKGLINDPD++ +FNIN                DW L  E H+++RR+ +PDE  FFPD + ++ LP I+Y  ILAPY+ EV+ ND+IKK+
Sc   76  CSIHDLEAAQEYALRLKKLSDELKGDLSIIMRAYLEKPRTTVGWKGLINDPDVNNTFNIN  135   Sc   61  DWALSNLEIAHSRIRRFESPDETPFFPDKIQKSFLPSINYPQILAPYAPEVNYNDKIKKV  120

Yl  128  KGLRLSRKVFCDLTDLGLPIASEMLDTISPQFIADLLSLGAIGARTTESQLHRELASGLS  187   Yl  119  YTDDIVPLVCAGTGDQPENYGSVMVCDIETLQALSRRIHFGKFVAESKFLSETERFTELI  178
         KGL+ +R++F +LT++GLPI SEMLDTISPQ+IADL+S GAIGARTTESQLHRELASGLS                Y + I+PL+   GD  N+GSV  DIE LQ+LSRRIHFGKFVAE+KF S+   +T+LI
Sc  136  KGLQSARQLFVNLTNIGLPIGSEMLDTISPQYLADLVSFGAIGARTTESQLHRELASGLS  195   Sc  121  YIEKIIPLISKRDGDDKNNFGSVATRDIECLQSLSRRIHFGKFVAEAKFQSDIPLYTKLI  180

Yl  188  FPVGFKNGTDGTLGVAVDAVQAASHPHHFMGVTKQGVAAITTTKGNENCFIILRGGKKGT  247   Yl  179  KNKDIAGIEAAITNSKVEETILARLGEKALAYGTDPTLRWSQRTQGKVDSEVVKRIYKEW  238
         FPVGFKNGTDGTL VAVDA QAA+H HHFMGVTK GVAAITTTKGNE+CF+ILRGGKKGT                K+KD+ GI  ITNS VEE IL RL +KA  YG DPT   +R  ++ E + +IYKE
Sc  196  FPVGFKNGTDGTLNVAVDACQAAAHSHHFMGVTKHGVAAITTTKGNEHCFVILRGGKKGT  255   Sc  181  KSKDVEGIMKNITNSAVEEKILERLTKKAEVYGVDPTNESGER---RITPEYLVKIYKEI  237

Yl  248  NYDAESVAECKK---ATESMLMVDCSHGNSNKDYRNQPKVSKAVAEQVAAGEKKIIGVMI  304   Yl  239  VIPLTKKVEVDYLLRRLE  256
         NYDA+SVAE K    A + LM+D SHGNSNKD+RNQPKV+  V EQ+A GE  I GVMI                VIP+TK+VEV+YLLRRLE
Sc  256  NYDAKSVAEAKAQLPAGSNGLMIDYSHGNSNKDFRNQPKVNDVVCEQIANGENAITGVMI  315   Sc  238  VIPITKEVEVEYLLRRLE  255

Yl  305  ESNIHEGNQKVPKEGPSALKYGVSITDACVSWETTVDMLTELANAVKERRNKN   357
         ESNI+EGNQ +P EG + LKYGVSITDAC+ WETT D+L +LA AV++RR N              Score:283 bits(725), Expect:7e-96,
Sc  316  ESNINEGNQGIPAEGKAGLKYGVSITDACIGWETTEDVLRKLAAAVRQRREVN   368            Method:Compositional matrix adjust.,
                                                                                Identities:143/258(55%), Positives:184/258(71%), Gaps:5/258(1%)

Score:536 bits(1381), Expect:0.0,
Method:Compositional matrix adjust.,
Identities:252/353(71%), Positives:300/353(84%), Gaps:3/353(0%)
```

Figure 25. Comparison of amino acid sequences of *S. cerevisiae* and *Y. lipolytica* Aro4 and Aro7 enzymes

Sequences of *Y. lipolytica* Aro4 and Aro7 enzymes blasted against the equivalent *S. cerevisiae* enzymes. In yellow are highlighted the positions usually mutated in *S. cerevisiae* to construct the feedback insensitive forms of the enzymes. In green are highlighted the conserved equivalent positions for *Y. lipolytica*. In Aro7 sequences the blue marks correspond to other sites that were also proposed as involved in regulation of the enzyme.

The activity of *Y. lipolytica* Aro4 and Aro7 enzymes, wild-type and mutated forms, were tested *in vivo* by measuring the Ehrlich products in the culture supernatant. As previously mentioned, all the strains have also an extra copy of *ARO8* and *ARO10* to increase the detectable amount of Ehrlich metabolites. As seen in Figure 26, the individual expression of each enzyme increases the flux through the pathway compared to the shikimate wild-type pathway strain (WT in Figure 26). Aro4^{K221L} showed the strongest effect, increasing the metabolites production more than six times compared to "wild-type"-strain. The higher effect of Aro4^{K221L} compared to the wild-type Aro4 form, suggest that the introduced mutation may have turn the enzyme into a feedback insensitive form. On the other hand, the activity of Aro7 does not change between the wild-type and the mutated form, hence, the enzyme was not deregulated by the mutation,

implying that regulation is not as described for this enzyme in *S. cerevisiae*. These results are in accordance with the enzymatic test done with the wild-type forms of the enzymes.

The enzymes were also tested together, after assembling them by GG. Contrary to what was expected, these joint expressions produced a lower amount of Ehrlich metabolites than the individual expressions. The negative effect was even stronger when the mutated forms were used (Figure 26). Furthermore, these strains grow around three times slower than the other engineered strains. A possible explanation is the appearance of some level of Tyr auxotrophy due to a stronger pull of the chorismate to the prephenate caused by Aro7, causing lower growth rate, thus, less Ehrlich metabolites. This should be further studied.
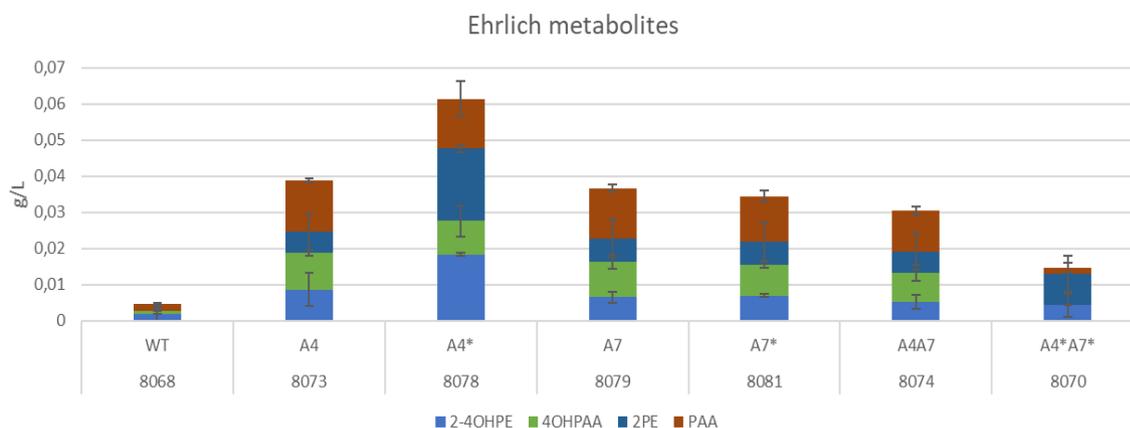
Figure 26. Evaluation of *Y. lipolytica* Aro4 and Aro7 activity by measuring Ehrlich metabolites production.

Strains expressing wild-type and mutated forms of the *Y. lipolytica* enzymes were cultured and then evaluated by the Ehrlich metabolites detected in the supernatants. Aro4* seems to be deregulated and, thus, produce more metabolites. On the contrary Aro7* does not present any difference with the wild-type form of the enzyme (A7). A: Aro gene; *: mutated version of the enzyme. Mutations on Aro4 (ylAro4K221L) and Aro7 (ylAro7G139S) are equivalent to feedback inhibited forms from *S. cerevisiae* (scAro4K229L and scAro7G141S). WT: shikimate wild-type pathway strain. The numbers under the genotype indicate their collection number (JMYnumber). All the strains are overexpressing *ARO8* and *ARO10*, including the WT.

These enzymes activities were also tested in a context of higher flux, in a genetic background of Aro1 and Aro2 overexpressed (Figure 27). The same profile was obtained, Aro4^{K221L} was the most active of the four enzymes forms tested and the joint expression of Aro4 and Aro7 (wild-types or mutated variants) has a negative effect. This confirms the regulation phenotypes identified for these enzymes. In addition, it can be observed that the overexpression of Aro1-Aro2 double the amount of metabolites detected (Figure 26 and Figure 27).
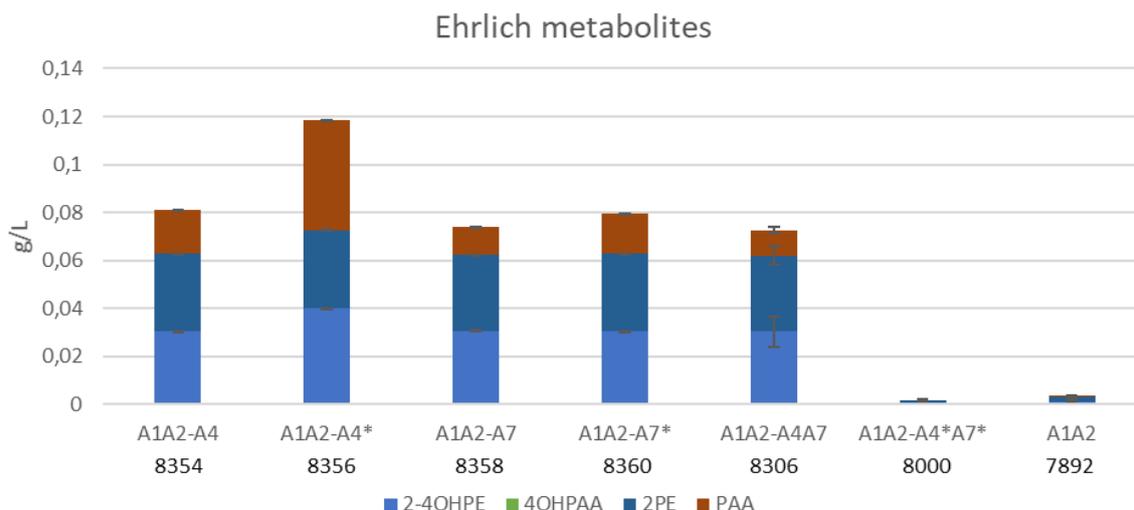
Figure 27. Evaluation of *Y. lipolytica* Aro4 and Aro7 activity by measuring Ehrlich metabolites production in strains overexpressing *Aro1* and *Aro2*.

The engineered strains are cultured, and then evaluated for the Ehrlich metabolites detected in the supernatant. Aro4* seems to be deregulated and thus produced more metabolites. On the contrary A7* does not have any difference with the wild-type form of the enzyme. A: Aro gene; *: mutated version of the enzyme (Mutations on Aro4 (ylAro4K221L) and Aro7 (ylAro7G139S) are equivalent to feedback inhibited forms from *S. cerevisiae* (scAro4K229L and scAro7G141S). The numbers under the genotype indicate their collection number (JMYnumber), they can be found in the strain table, in the annexe part, to see the complete genotype of the strain.

Due to the particularly low production of metabolites by the strain JMY8000 (Figure 27), the gene expression was analysed by qRT-PCR. Results showed that the four genes (*ARO1, ARO2, ARO4^{K221L}* and *ARO7^{G139S}*) are being overexpressed (Figure 28), indicating that the absence of Ehrlich metabolites is not caused by lack of enzyme expression. The causes of this effect are still unclear and should be further studied.
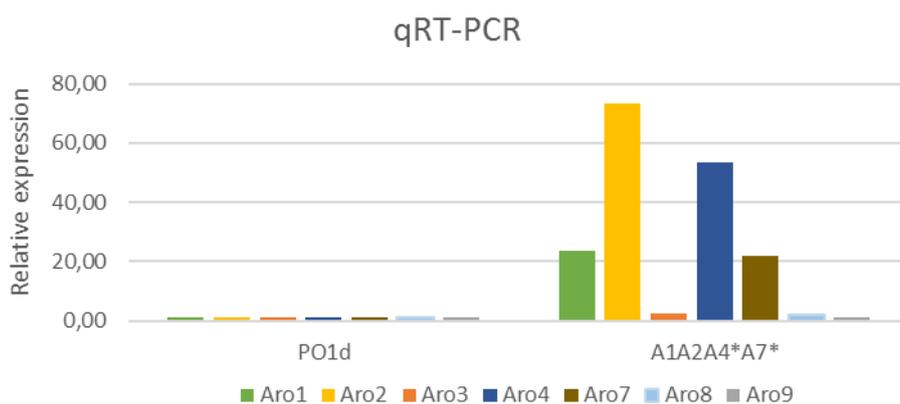
Figure 28. qRT-PCR analysis

Level expression of genes in an engineered strain compared to the non-engineered strain. This shows that the genes incorporated into the genome of *Y. lipolytica* are overexpressed. A: Aro gene; *: mutated version of the enzyme

Due to the fact that the co-expression of $ARO4^{K221L}$ and $ARO7^{G139S}$ has a negative effect on the production of Ehrlich metabolites, underlining possible unknown regulations, they were not further used for the construction of AAA chassis strain.

3.2.3   AAA chassis strain as cell factory to produce valuable compounds

The strain constructed in this work and identified as the best AAA producer strain was selected as the AAA chassis strain. This strain, JMY8175, overexpressing *ARO1-ARO2*, *scARO4^{K229L}-scARO7^{T226I}* and *scARO3^{K222L}* was used as a cell factory for the biosynthesis of high-value products, with a wide range of applications. Hereafter are described the results obtained in this regard.

3.2.3.1   2-Phenylethanol

2PE is a metabolite obtained from the catabolism of Phe through Ehrlich pathway (Figure 6), characterized by its rose-odour like aroma. Thanks to this pleasant scent, this molecule is intensively used in the cosmetic and food industries, and its demands increases through the years as a natural flavour compound.

Phe and Tyr can both be assimilated by the Ehrlich pathway. The partition among the different derived metabolites depend on enzymes activities involved in prephenate metabolism, that determine the amount of Phenyl-pyruvate or 4-OH-Phenyl-pyruvate, and on the redox state of the cell that determine the ratio of alcoholic to acid form of the phenyl-pyruvate-derived compound (Hazelwood et al., 2008) (Figure 6). With this in mind, one can suppose that the balance among the four Ehrlich metabolites can be easily changed, favouring the production of one of the compounds against the others.

As mentioned in section 3.2.1.2, the production of Ehrlich metabolites at detectable levels by the *Y. lipolytica* strain used in this work needs the overexpression of *ARO8* and *ARO10*. On that same section, I showed the increased production of these metabolites in strains engineered for higher flux through the AAA pathway (Figure 20 and Figure 21). Thus, the constructed chassis strain (*ARO1-ARO2*, *scARO^{4K229L}-scARO7^{T226I}* and *scARO3^{K222L}*), overexpressing *ARO8* and *ARO10* (JMY8109), is a good platform to produce Ehrlich metabolites and consequently 2PE (Figure 21).

After 10 days of culture on 50 mL YNBglc40, the strain JMY8109 produces 0.110 g/L of 2PE, while a strain only overexpressing *ARO8* and *ARO10* but without any modification of the shikimate pathway (JMY7891) produces half the amount of 2PE (0.0476 g/L). Additionally, the chassis strain produces three times more of total Ehrlich metabolites than the wild-type shikimate pathway strain (Figure 29), meaning that it is potentially feasible to increase the difference of 2PE production between these two strains by changing the balance of these metabolites.
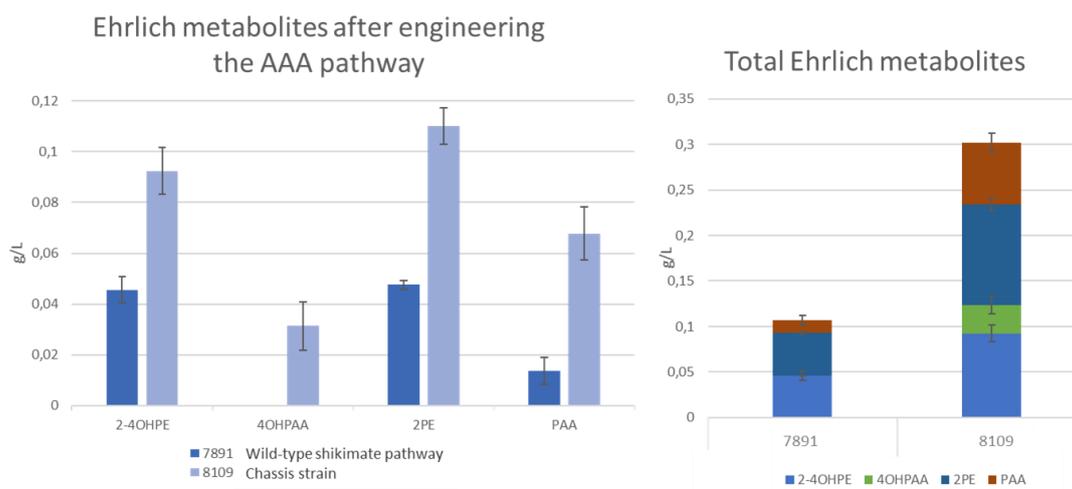
Figure 29. Ehrlich metabolites produced in engineered and non-engineered strains.
The chassis strain JMY8109 and the shikimate wild-type-pathway strain JMY7891 were cultured for 10 days on YNBGlc40. Left graph shows the production of individual Ehrlich metabolite, while right graph shows the accumulated amount of Ehrlich metabolites. Both strains overexpress *ARO8* and *ARO10*. YNB: minimal Yeast Nitrogen Based medium. Glc40: 40 g/L glucose.

In an attempt to increase the production of these metabolites, and consequently of 2PE, different amounts and types of carbon sources were evaluated. Therefore, two of the AAA producer strains of this work, JMY8032 and JMY8109, which differ in the overexpression of *TKL* and *ARO3^{K222L}* (see Figure 26), were tested in 10 and 40 g/L of glucose and in 40 g/L of glycerol (Figure 30). In medium containing 40 g/L of glucose the strains produced three times more Ehrlich metabolites compared to the production in medium containing only 10 g/L of the same sugar. The production of 2PE, particularly, increased 40 times in these conditions, from 0.0025 to 0.095 g/L (almost the same values were obtained from both strains). The production of the other metabolites was also increased, but only two times. The use of glycerol as carbon source, on its part, showed a positive impact on the production of Ehrlich metabolites compared to glucose.

The production of 2PE was doubled while the Tyr-derived metabolites (2-4OHPE and 4OHPAA) production only increased around 1%. The production of PAA, on the other hand, increased 50% in the strain JMY8032 and 150% in JMY8109. Except in this latter case, the increase in the production of metabolites had the same profile in both strains (Figure 30). This test evidenced that the amount, as well as the type, of carbon source has a big impact on the production of these metabolites. Even though the production of 2PE was almost the same between both strains, JMY8109 showed a higher production of total Ehrlich metabolites compared to JMY8032, in all the condition tested, mostly due to the production of PAA.
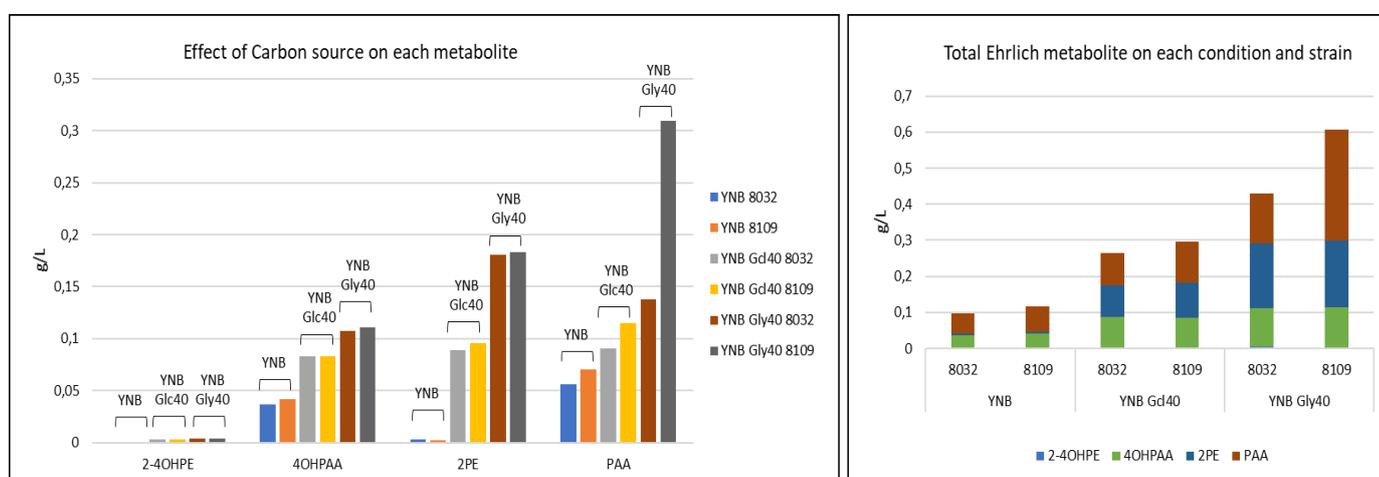
Figure 30. Effect of carbon source on the production of Ehrlich metabolites.

Two AAA engineered strains, overexpressing *Aro1-Aro2-scAro4^{K229L}-scAro7^{T226I}*, were tested with different amount and type of carbon sources. The strain JMY8032 overexpress in addition *TKL* while the strain JMY8109 express the deregulated form of Aro3 (ARO3^{K222L}). Left graph shows the production of individual Ehrlich metabolite in each media tested and for both strains. Right graph shows the accumulated amount of total Ehrlich metabolites for each condition and strain. Both strains overexpress *ARO8* and *ARO10*. YNB: minimal Yeast Nitrogen Based medium. Glc40: 40 g/L glucose. Gly40: 40 g/L glycerol

The overexpression of *TKL1* has for objective to increase the availability of the precursor E4P through the PPP. In addition, the use of glycerol as carbon source was described to increase the production of erythritol, which also uses E4P as precursor. Furthermore, the production of erythritol was increased when *TKL1* was overexpressed and glycerol used as carbon source (Carly et al., 2017b). Thus, the use of glycerol as carbon source for a strain overexpressing Tkl1, as it is the case for JMY8032, was expected to increase the production of Ehrlich metabolites. However, the increase observed was lower than in the strain overexpressing scARO3^{K222L} (JMY8109) (Figure 30). This could be due to the mentioned higher activity of Aro3 compared to Aro4 (Hartmann et al., 2003), which can be responsible of a stronger pull of precursors into

the shikimate pathway. Therefore, the use of glycerol as carbon source for the strain JMY8054 which overexpress Tkl1 and scAro3^{K222L} (Figure 21) has to be tested. This may increase the availability of E4P and at the same time force it to enter the shikimate pathway more efficiently, thus decreasing its use in competitive pathways.

Another factor that can limit the high-yield production of 2PE in microbial cell factories is its cytotoxic activity on microorganisms (Etschmann et al., 2002; Hassing et al., 2019; Hua and Xu, 2011). One possible approach to avoid this toxic effect is to remove the compound from the aqueous cell phase, which is possible due to 2PE solubility on organic solvents. This *in situ* product removal proved to be efficient to increase 2PE yields production by other microorganisms (Etschmann and Schrader, 2006). Thus, in order to increase the production by *Y. lipolytica*, 10% decane was added into the culture medium of the JMY8109 strain. This approach led to a 50% increase in the production of 2PE, as well as in the total Ehrlich metabolites (Figure 31).
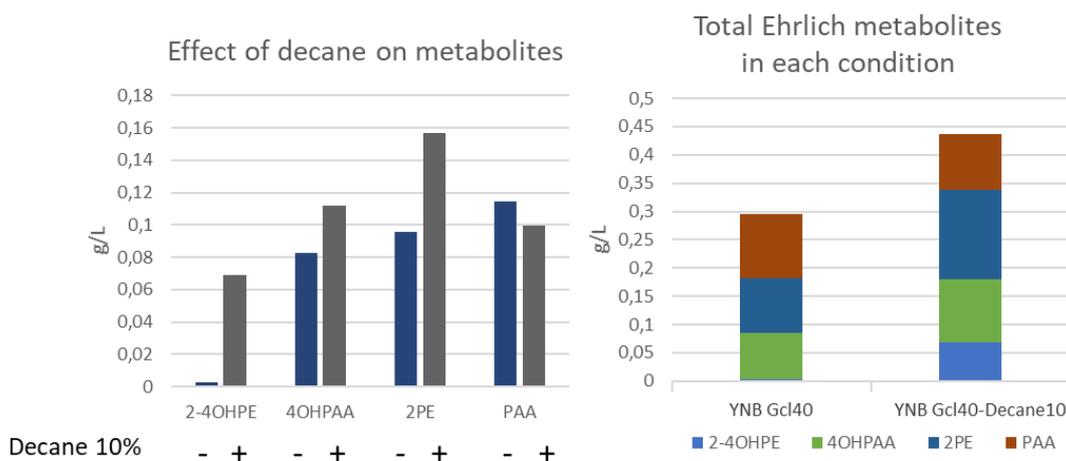
Figure 31. Effect of decane on Ehrlich metabolites production.

The chassis strain JMY8109 was cultured with and without decane in the medium. Left graph shows the impact of decane on the production of individual Ehrlich metabolites, while right graph shows the accumulated amount of Ehrlich metabolites in each condition. YNB: minimal Yeast Nitrogen Based medium. Glc40: 40 g/L glucose. Decane10: 10% decane

Additionally, to evaluate if the production of 2PE was being limited by a still constrained flux through the AAA pathway, the chassis strain (JMY8109) and two strains without modifications on the shikimate pathway (W29 (Wild type) and JMY7891 ("Wild type" + Aro8Aro10)) were cultured for five days in minimal media supplemented with Phe or Tyr (1 g/L). The addition of Phe increased the production of Ehrlich metabolites in around 0.25 g/L in the three strains while the addition of Tyr produced

an increase of around 0.1 g/L in strains JMY7891 and JMY8109 (Figure 32). This in-dicates that the production of 2PE is still limited by the shikimate pathway and not due to a saturation of the Ehrlich pathway. Thus, the chassis strain can still be improved. The reason for the lack of response of W29 to the addition of Tyr is unknown. The consumption of Tyr was verified by HPLC, as well as for the other strains, so it may have been metabolized by other pathway as this strain is not overexpressing the Aro8 and Aro10 enzymes that favors the Ehrlich pathway.
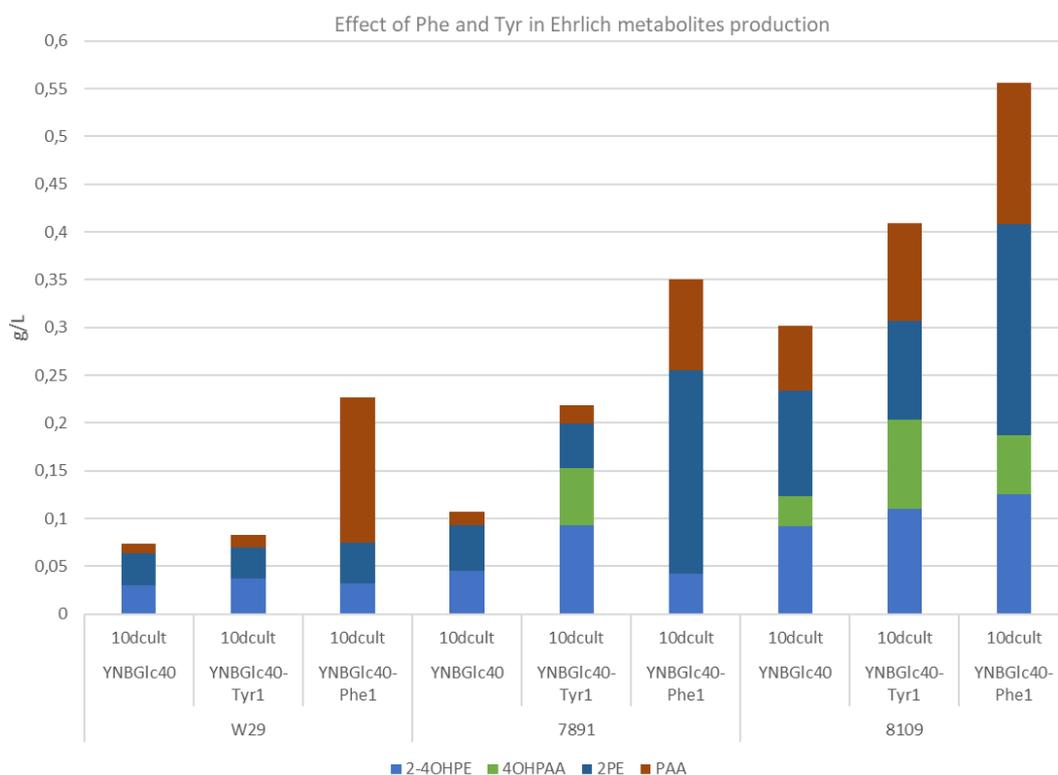
Figure 32. Effect of supplementing the medium with Phe and Tyr on the Ehrlich metabolites production.

W29: *Y. lipolytica* wild-type strain. 7891: JMY7891 ("WT" + Aro8Aro10). 8109: JMY8109 (Chassis strain +Aro8Aro10) YNB: minimal Yeast Nitrogen Based medium. Glc40: 40 g/L glucose. Phe1: 1 g/L Phenylalanine. Tyr1: 1g/L tyrosine

In these bioconversion-tested conditions, the JMY8109 strain produced higher amount of total Ehrlich metabolites than the strains with wild-type shikimate pathway. However, the addition of Phe or Tyr had a bigger impact on Ehrlich metabolites pro-duction on the strains with not engineered shikimate pathway than on the chassis strain (Figure 32). The Phe-derived compounds (2PE and PAA) produced by W29 and JMY7891 increased five times after the addition of 1 g/L of Phe, while JMY8109 in-creased its production twice. On the other hand, the addition of Tyr did not have any effect on W29 but increased three and two times the tyrosine-derived products (2-

4OHPE and 4OHPAA) produced by JMY7891 and JMY8109, respectively. This more limited bioconversion observed by the engineered strain may be due, for instance, to regulation mechanisms or to toxicity of the produced metabolites.

Furthermore, the addition of Phe increases the production of all the Ehrlich metabolites almost in the same proportions in the strain JMY8109. On the contrary, the effect of Tyr was almost limited to Tyr-derived products. Chemically, the difference between Phe and Tyr is the *para*-hydroxyl group present in this last one (Figure 6). In higher organisms, the enzyme phenylalanine hydroxylase is responsible for the direct conversion of Phe into Tyr by hydroxylation. This biological conversion is described as an irreversible process (Matthews, 2007; Mosst and Schoenheimer, 1940). Even though this enzyme has not been described in yeasts (Wang et al., 2013), the hydroxylation of the supplemented Phe may occur, explaining the concomitant increase of Phe and Tyr derived products under this condition. On the contrary, as the hydrolysis of the hydroxyl in Tyr may not occur naturally, or at a very low rate the supplemented Tyr almost do not influence the Phe-derived Ehrlich metabolites.

Up to now, the higher amount of *de novo* 2PE production reached during this work was of 0.18 g/L using glycerol as carbon source. Under this condition, 0.6 g/L of total Ehrlich metabolites were produced, with 0.49 g/L being from Phe-derived products (Figure 30). Thus, the amount of 2PE produced can be increased by optimizing culture conditions to favor the ethanol over the acid form and by expressing a dehydrogenase. Furthermore, the shikimate pathway can still be improved without saturation of the 2PE production.

3.2.3.2 Naringenin and Resveratrol

Naringenin and resveratrol are secondary plant metabolites derived from AAAs with potential therapeutic and nutritional uses for humans. Due to the limited plant productions yield, their biosynthesis through microbial cell factories using an AAA optimized chassis strain is of interest. As introduced before, *Y. lipolytica* has intrinsic properties that make it particularly interesting for its development as a cell factory for the production of this compounds: (i) its hydrophobic environment, which is critical for regioselectivity and stereoselectivity in hydroxylation by P450 enzymes involved in these biosynthetic pathways (M. Gao et al., 2017; Nthangeni et al., 2004; Suástegui and

Shao, 2016), and (ii) the high content of malonyl-CoA (Christen and Sauer, 2011), which is highly required for the synthesis of these compounds; three molecules of malonyl-CoA are needed per molecule of naringenin or resveratrol produced (Figure 33). Additionally, the biosynthetic pathways of these compounds are very similar (Figure 33), diverging only after coumaroyl-CoA production, which ease the use of engineered strains and reduces the number of different heterologous genes needed for their production.

Heterologous pathway in *Y. lipolytica*
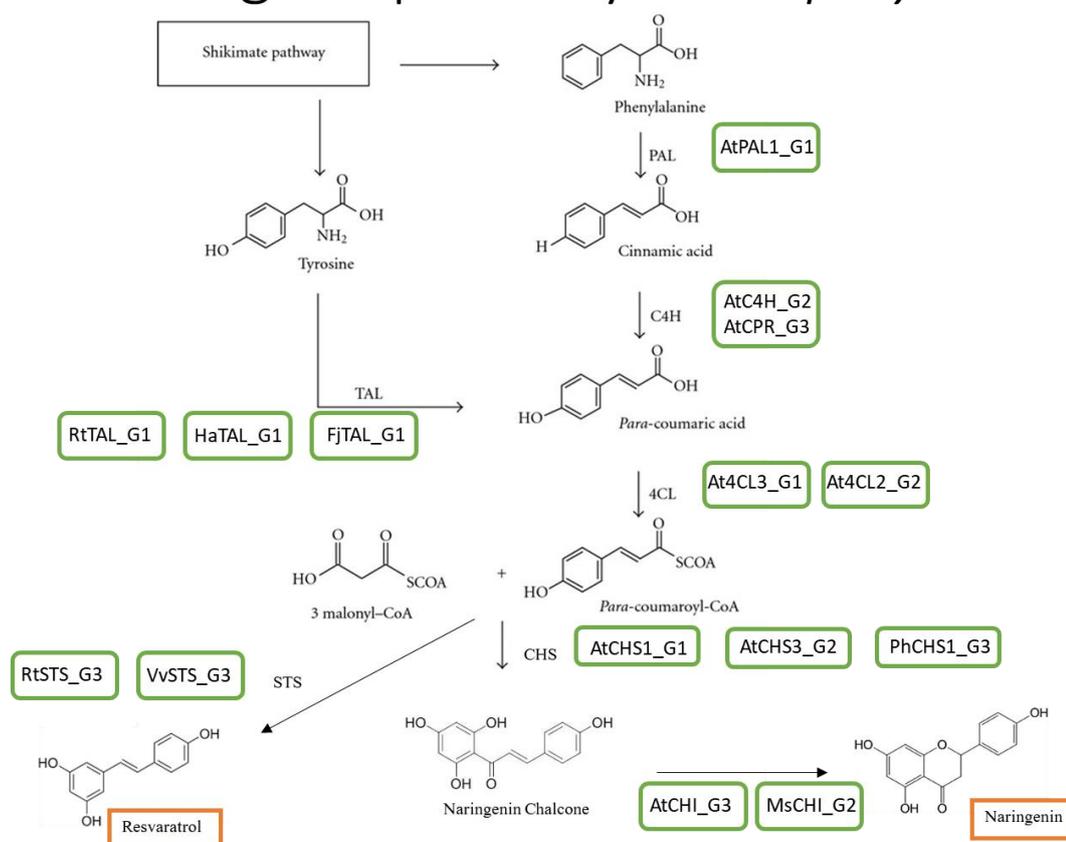
Figure 33. Heterologous pathway for the biosynthesis of naringenin and resveratrol in *Y. lipolytica*.

In green squares are shown the heterologous codon-optimized genes, synthetised for GG assembly and available for their test in *Y. lipolytica*. The G-number indicates the GG position for which the genes has been constructed.

It has been shown that the selection of the ortholog genes, to be expressed as in the heterologous metabolic pathway constructed, plays a tremendous role in the final titer of flavonoids and stilbenes (Xu et al., 2013). As a result, different combination of orthologs genes and promoters need to be explored to identify the best combination that maximizes the biosynthesis of this kind of compounds. With this aim, a large number

of genes from different organisms were codon optimized for their expression in *Y. lipolytica* and designed according to our GG assembly technique (Figure 33 and Table 4).

The constructed AAA chassis strain, identified as the best AAA producer in my

Table 4. List of enzymes from different organisms involved in the biosynthesis of naringenin and/or resveratrol

Plasmid name	Organism	Enzyme name	EC number	Reference
AtPAL1	Arabidopsis thaliana	phenylalanine ammonia-lyase	4.3.1.24	Koopman *et al.*, 2012
AtC4H	Arabidopsis thaliana	Cinnamate 4-hydroxylase	1.14.14.91	Koopman *et al.*, 2012
AtCPR1	Arabidopsis thaliana	cytochrome P450 reductase	1.6.2.4	Koopman *et al.*, 2012
At4CL3	Arabidopsis thaliana	4-coumaric acid-CoA ligase	6.2.1.12	Koopman *et al.*, 2012
At4CL2	Arabidopsis thaliana	4-coumaric acid-CoA ligase	6.2.1.12	Koopman *et al.*, 2012
AtCHI1	Arabidopsis thaliana	Chalcone isomerase	5.5.1.6	Koopman *et al.*, 2012
AtCHS3	Arabidopsis thaliana	chalcone synthase	2.3.1.74	Koopman *et al.*, 2012
AtCHS1	Arabidopsis thaliana	chalcone synthase	2.3.1.74	Koopman *et al.*, 2012
RtTAL	Rhodosporidium toruloides	Phe/Tyr ammonia-lyase	4.3.1.23/24/25	Lv *et al.*, 2019
FjTAL	Flavobacterium johnsoniae	Tyrosine ammonia-lyase	4.3.1.23	Jendresen *et al.*, 2015
HaTAL	Herpetosiphon aurantiacus	Tyrosine ammonia-lyase	4.3.1.23	Li *et al.*, 2015
PhCHS	Petunia x hybrida	chalcone synthase	2.3.1.74	Lv *et al.*, 2019
MsCHI	Medicago sativa	Chalcone isomerase	5.5.1.6	Lv *et al.*, 2019
CrCPR	Catharanthus roseus	cytochrome P450 reductase	1.6.2.4	Lv *et al.*, 2019
VvSTS1	Vitis vinifera	Stilbene synthase	2.3.1.95	Li *et al.*, 2015
RtSTS	Rheum tataricum	Stilbene synthase	2.3.1.95	Forster *et al.*, 2006

series of AAA engineered strains ($JMY8175 = ARO1\text{-}ARO2$, $scARO4^{K229L}\text{-}scARO7^{T226I}$ and $scARO3^{K222L}$), was used to plug the heterologous pathway. Nevertheless, in order to compare the production in differentially engineered strains and aiming to assert the efficacy of the chassis strain, other intermediate strains of the chassis construction series were used. On the contrary to what was previously done during this work, the strains where the heterologous pathways were plugged do not overexpress *ARO8-ARO10* cassette, thus avoiding the catabolism of Phe and Tyr by the Ehrlich pathway.

In order to evaluate the effect of the different heterologous genes in the production of naringenin and resveratrol, several constructions were done. As an example, constructions using AtPAL or RtTAL were done in order to evaluate the difference, if any, on using Phe or Tyr as precursor (Figure 33). The construction of different strains with ortholog enzymes aimed to evaluate their different influence in the studied biosynthesis. Also, strains bearing several copies of one or more genes of the pathways were constructed in order to improve the production. Regarding the naringenin pathway expression, for instance, it was described that the rate limiting step was the CHS enzyme. In an *in vitro* study, the best ratio of 4CL/CHS/CHI was determined as 10:10:1, and malonyl-CoA was the limiting factor (Zang et al., 2019). The more relevant strains constructed are shown in Table 5, also indicating the amount of product detected. The complete list is in the Annexes section (6.1.3). Constructed strains were flask cultured on YNBglc40 and supernatants were then analysed by HPLC. Unfortunately, very few of the tested strains produced some detectable amount of naringenin or resveratrol.

Regarding naringenin, differences in the amount produced are evidenced depending on the genetic background strain. It can be observed that naringenin is detected in strains with engineered AAA pathway, while no naringenin is produced in strains where this pathway was not modified (e.g. strains JMY8247 and JMY8249). Additionally, the increase on the number of copies of some of the genes increase the production of this compound (e.g. strains JMY8247 and JMY8290). On the other hand, some constructions did not lead to detectable levels of compound; this should be further analysed in order to identify the cause.

The strain JMY8247 (*ARO1-ARO2- scARO4^{K229L}-scARO7^{T226I} + RtTAL-At4CL3-PhCHS*x2-*MsCHI*) was the first from which production of naringenin was detected, thus used to do complementary experiments. As indicated in its genotype, this strain has an engineered AAA pathway and the naringenin pathway contains two copies of the gene *PhCHS* (x2), which was identified as being the bottleneck of the pathway. In minimal medium (YNB) after six days of culture, this strain produces 1,5 mg/L of naringenin. In order to verify this result, two approaches were used. First, 1,5 mL of supernatant were concentrated, using a vacuum concentrator apparatus (SpeedVac from Thermo), and then measured again by HPLC. This resulted in a bigger and clearer HPLC signal, corresponding to 2,1 mg/L naringenin in the original sample, which is in

accordance with what was detected in the first HPLC run. In addition, the presence and concentration of naringenin was confirmed using a biosensor system developed by

Table 5. List of the most significant strains constructed for the production of naringenin and resveratrol

	Naringenin	mg/L
	AtPal	
JMY8030	JMY195 + atPAL-atC4H-CPR + atCHSI-at4CL2-atCHI1	0
JMY8027	JMYY195 + scAro4K229L-scAro7T226I + atPAL-atC4H-CPR + atCHSI-at4CL2-atCHI1	0
JMY8028	JMY195 + Aro1-Aro2_Ura + scAro4K229L-scAro7T226I + atPAL-atC4H-CPR_Nat + atCHSI-at4CL2-atCHI1_hph	0
	RtTAL	
JMY8198	JMY195 + RtTal-MsCHI-PhCHS	0
JMY8249	JMY195 + RtTal-MsCHI-PhCHS **x2** + At4Cl3	0
JMY8227	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + RtTal-MsCHI-PhCHS + At4Cl3	0
JMY8247	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + RtTal-MsCHI-PhCHS**x2** +At4Cl3	3,4296
JMY8294	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + RtTal-MsCHI-PhCHS**x3** +At4Cl3**x2**	0
JMY8290	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + RtTal**x2**-MsCHI**x2**-PhCHS**x3** +At4Cl3	11,968
JMY8293	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + RtTal-MsCHI-PhCHS**x2** +At4Cl3 + atPAL-atC4H-CPR	0
JMY8200	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + scAro3K222L + RtTal-MsCHI-PhCHS	0
JMY8230	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + scAro3K222L + RtTal-MsCHI-PhCHS + 4Cl3	0
JMY8284	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + scAro3K222L + RtTal-MsCHI-PhCHS**x2** + At4Cl3**x2**-At4Cl2	8,024
JMY8286	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + scAro3K222L + RtTal**x2**-MsCHI**x2**-PhCHS**x3** + At4Cl3	10,6352
	RESVERATROL	
JMY7969	JMY195 + HaTAL-at4CL2-vvSTS	0
JMY8213	JMY195 + HaTal-At4CL2-VvSTS + At4Cl3	0,34
JMY8196	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + HaTal-At4Cl-VvSTS	0,36
JMY8215	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + HaTal-At4CL2-VvSTS + At4CL3	0,66
JMY8192	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + scAro3K222L + HaTal-At4CL2-VvSTS	0,31
JMY8214	JMY195 + Aro1-Aro2 + scAro4K229L-scAro7T226I + scAro3K222L + HaTal-At4CL2-VvSTS + At4CL3	0,55

The table is divided in naringenin and resveratrol producer strains. Black lines divide group of strains with different AAA pathway background. Numbers in red indicate how many times the gene has been introduced into the genome of the strain. In naringenin list a division is also done between constructions containing PAL or TAL. JMY195 is the "wild-type" strain. + are used to separate the different expression cassettes used. Genes in green are the ones, from the engineered AAA pathway, that were confirmed by qRT-PCR to be present in the strain (see the text and Figure 35).

Trabelsi and co-workers. This TF-based biosensor express, in *E. coli*, a red fluorescent protein in the presence of naringenin (Trabelsi et al., 2018). For the study of *Y. lipolytica* production strain, the functionality of the biosensor was first validated in our hands. To this end, the biosensor was tested with a commercial standard of naringenin and a calibration curve was established. In a second time, tests were done to set the conditions for co-culture detection. For the evaluation of *Y. lipolytica* strains, 100 µl of an overnight preculture of *E. coli* containing the biosensor system was added into a 96-well plate together with 100 µl of the *Y. lipolytica* culture to evaluate. The mixture was incubated for 5 h at 37°C with constant shake in a microtiter plate reader, and fluorescence was measured every 15 minutes. Using this biosensor approach the concentration of naringenin detected in the sample was of 1,25 mg/L, which is also in accordance with the value obtained by HPLC, validating the sensitivity and accuracy of the biosensor system for future screening of *Y. lipolytica* production strains.

Afterwards, the production of naringenin was tested by culturing the strain in different media. Like the production of 2PE, the production of naringenin increase with the amount of glucose used in the media, and glycerol proved to be a better carbon source than glucose (Figure 34). The amount of naringenin detected was three times higher in a medium containing 40 g/L of glucose than in a medium with only 10 g/L of glucose. The use of glycerol as carbon source, on the other side, increase the production
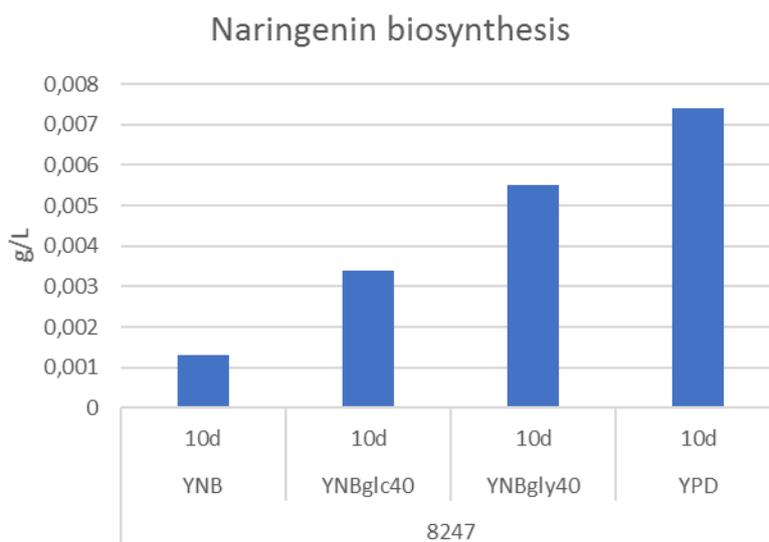
Figure 34. Influence of the media in naringenin production.

8247: strain JMY8247. Gcl40: glucose 40g/L. Gly40: glycerol 40g/L. 10d: 10 days culture

of naringenin of about 1.5 times. The highest production of naringenin, 7 mg/L, was achieved using the rich YPD medium (Figure 34).

Regarding resveratrol, some of the constructed strains showed some production, as shown in Table 5. Even though the detected amounts are very low (between 0.2-0.65 mg/L), it can be notice that the amount increase when an extra copy of the gene *4CL* is included, meaning that the expression or activities of the enzymes of the pathway are a bottleneck. In addition, the strains with engineered AAA pathway are able to produce higher amounts of resveratrol than the shikimate wild-type pathway strain (Table 5).

Even though some strains produced detectable levels of naringenin or resveratrol, the amounts obtained were low and not in accordance with what was expected nor with what was published by other groups working in *Y. lipolytica* (Lv et al., 2019b, 2019a; Palmer et al., 2020) where they achieved higher production rates, with lower times of culture and using the same heterologous genes that in this work and with a much less engineered AAA pathway. Several hypotheses can explain the low level of production of a heterologous compound. For instance, low or absence of gene expression due to it locus insertion site or not optimal codon optimization, low enzyme activity and/or metabolic burden caused by the expression of heterologous genes.

In order to better understand the low production of our strains, the expression level of shikimate and heterologous pathway genes was analysed by qRT-PCR (Figure 35) for four constructions (two producing naringenin and two producing resveratrol). Several genes overexpressed in the AAA pathway were not expressed, meaning that the cells have lost at least part of the expression cassette introduced to improve the AAA pathway (Figure 35). Consequently, the AAAs pool, precursor of the heterologous pathways, was no longer increased. This may have occurred during the treatment with the recombinase Cre enzyme of the chassis strain to recover selection markers with the Cre-Lox system. The Cre enzyme excise the markers thanks to a recombination event between the lox sequences that are flanking it. Even though it is a rare event, the recombination loop can cause the loss of bigger parts of the cassette or even of the entire cassette.

On the other hand, although the heterologous genes involved in the biosynthesis of naringenin and resveratrol are being expressed in the cell, most of them are expressed

at levels lower than the actin, indicating a relatively low expression (Figure 35). This low expression can be due to the gene sequences that were used; even though they were codon optimized it may be necessary to change the algorithm and do a new optimization. Also, the locus in the genome where the cassette was inserted can have an influence on the level expression of the cassette (Bordes et al., 2007; Leplat et al., 2015). This may be the cause of the difference observed on the expression level of the cassettes between the strains, JMY8247 and JMY8230, and JMY8213 and JMY8217 (Figure 35-A and B Fold Change). This different expression level highlights the importance of screening several clones of the same construction in order to identify the best producer. Also, high expression locus can be identified and used systematically to insert expression cassettes in order to avoid this locus-related problem. This qRT-PCR evaluation also allows us to evidence possible bottleneck, thus the expression level of At4CL was identified as a possible obstacle on the biosynthesis of naringenin and resveratrol. Furthermore, it can be observed the effect of increasing the number of copies of a gene. The expression of PhCHS in JMY8247, where two copies are present, is almost seven times higher than in the strain JMY8230 where only one copy is present.

Altogether, the loss of the engineered AAA pathway and a relatively low expression of the heterologous genes explain the very low, or lack of, heterologous product biosynthesis in the constructed strains.

Nonetheless, the strain producing naringenin or resveratrol and tested by qRT-PCR (JMY8247 and JMY8214) still have part of the engineered AAA pathway (Figure 35). Thus, expressing the feedback insensitive forms of scAro4 and scAro7 or scAro3 (*scARO4^{K229L}*, *scARO7^{G141S}* and *scARO3^{K222L}*) may increase the availability of Phe and Tyr sufficiently to allow the production of naringenin or resveratrol. With this encouraging result, further engineering is envisaged.

As a result, a new run of production strains should be constructed taking the previous results into consideration. In this sense, the AAA chassis strain should be transformed again with the heterologous pathways, however, the genotype has to be validated after the Cre treatment and before going to the next modification. In addition, the level of expression of heterologous genes from the naringenin and resveratrol pathway has to be increased, for instance, using stronger promoters, increasing the number of copies of the gene or expressing other ortholog genes.

Figure 35. Gene expression of engineered pathways in cell factory strains.

A qRT-PCR test was done in strains harbouring the heterologous biosynthetic pathways of naringenin (A) or resveratrol (B). The expression of actine (*ACT1*) is used to calculate the relative expression level, while fold change compares the expression level of the engineered strain towards the wild-type strain (Y195). The theoretical genotype of the strains is shown under their corresponding number (8247, 8230, 8213, 8214), which correspond to JMY-collection numbers.

3.2.3.3   Melanin

In the course of experiments, I observed after several days of culture of the AAA engineered strains, the apparition of a brown colour in the media. The colour intensity was not unified across the strains and it was clearly linked to their genotype. The engineered strains with the higher flux through the AAA pathway showed the more intense brown colour (Figure 36).
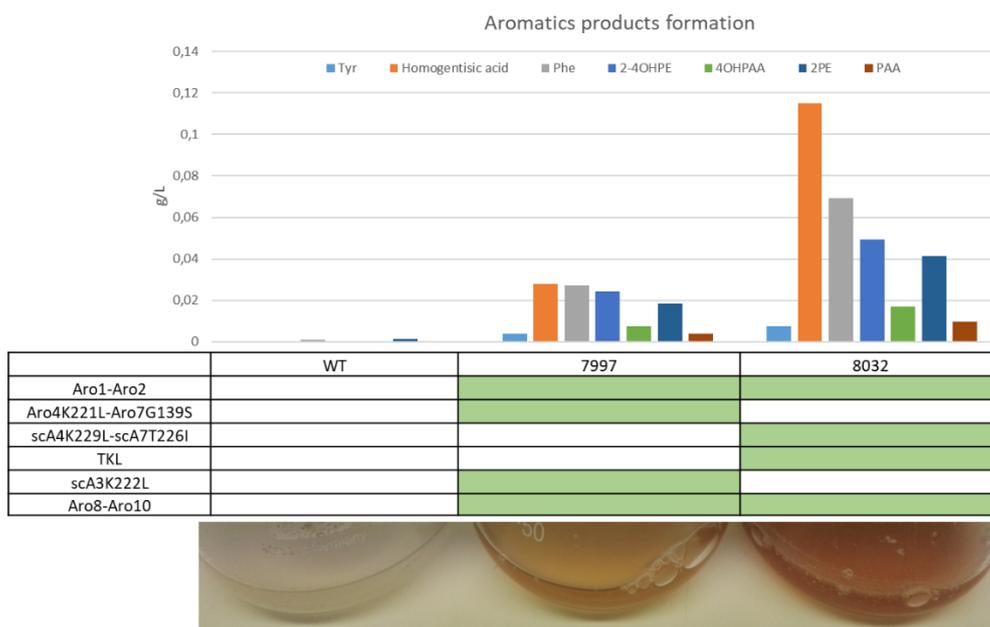
Figure 36. Brown pigment production on three different *Y. lipolytica* strains

The genotype of the strains (green filled cases, middle part of the figure), the amount of aromatic compounds detected on their culture supernatant (higher part of the figure) and the image of their differentially coloured media (lower part of the figure) are shown.

The production of brown pigments by *Y. lipolytica* was already reported, and described as a polymer composed of a core of aromatic residues derived from Tyr (Carreira et al., 2001a). This polymer, pyomelanin, is produced by autoxidation and polymerization of homogentisic acid (HGA) accumulated outside the cell (Turick et al., 2010) (Figure 37). Since the engineered strains constructed during this work are producing higher amounts of Phe and Tyr than *Y. lipolytica* wild type, one can expect that the brown pigment is related to the production of pyomelanin.
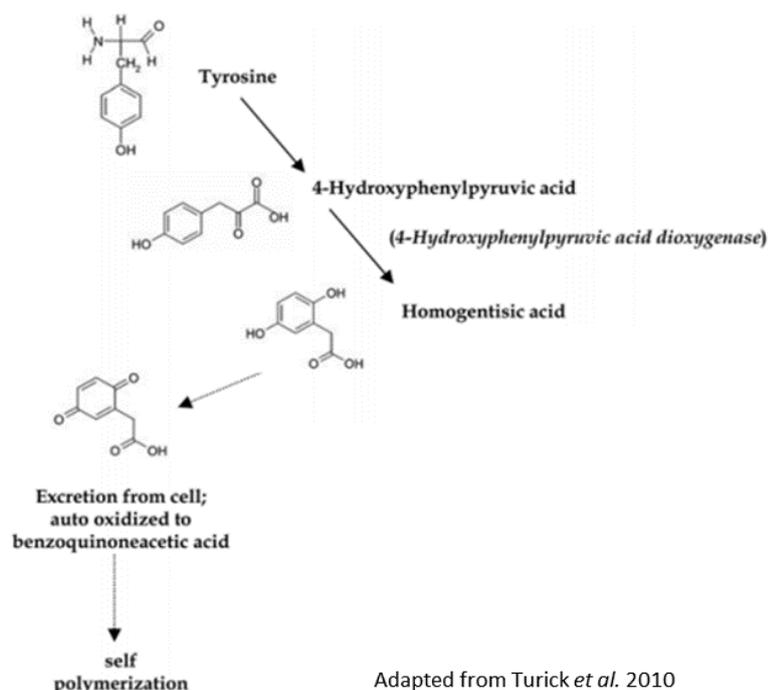
Adapted from Turick *et al.* 2010

Figure 37. Pyomelanin biosynthesis pathway.
Tyrosine is transformed into homogentisic acid, which is accumulated outside the cell where it is oxidized and polymerized to form pyomelanin.

In addition, the fact that the brown colour appears after several days of culture and that the intensity increases with time (Figure 38-A) is in accordance with the description of Carreira and co-workers. They observed that the pigment precursor was accumulated outside the cell during the exponential phase of growth and that pigment formation occurred during the stationary phase as a result from the oxidation of the precursor (Carreira et al., 2001a).

In order to confirm that the brown pigment observed in the engineered strains is indeed pyomelanin, different tests were conducted, and the results are discussed hereafter.

First, it was observed that the addition of Phe or Tyr in the culture medium accelerate the appearance of brown colour in the strains with engineered shikimate pathway. Tyr being the one that triggers the higher production (Figure 38-A Tyr and Phe 1g/L). The amount of Tyr could not be increased due its solubility, but when increasing the amount of Phe, the intensity of brown pigment increases and the time for its appearance decreases (Figure 38-A Phe1, Phe2 and Phe7). In the case of *Y. lipolytica* wild-type strain, the brown pigment started to appear only after 30 days of culture in YNB-Tyr1

medium (data not shown), differing from what was observed by Ben Tahar and co-workers where they describe the pyomelanin production on *Y. lipolytica* wild-type strain after five days culture in a medium containing 1g/L of Tyr (Ben Tahar et al., 2019). Brown pigments on *Y. lipolytica* wild-type strains was not seen in any other of the tested conditions. The presented results suggest that the pigment is derived from Phe and Tyr.

To further characterize the obtained pigment, ascorbic acid, which is described as an inhibitor of the synthesis of melanin due to its antioxidant properties (Ben-David et al., 2018; Yang et al., 2018), was added to the media. The brown pigment production was clearly reduced when 10 mM of ascorbic acid were added to the media, suggesting that the oxidation of the intermediate was being inhibited (Figure 38-B). In addition, when no buffer was added to the media, the brown pigment was not produced, even in the presence of AAA. This can be due to the natural acidification of the media during *Y. lipolytica* growth which can impair the oxidation of the intermediate (Figure 38-C). *Y. lipolytica* can naturally secrete several organic acids like citric acid, α-ketoglutaric acid, or pyruvic acid under several conditions (Finogenova et al., 2005; Otto et al.,
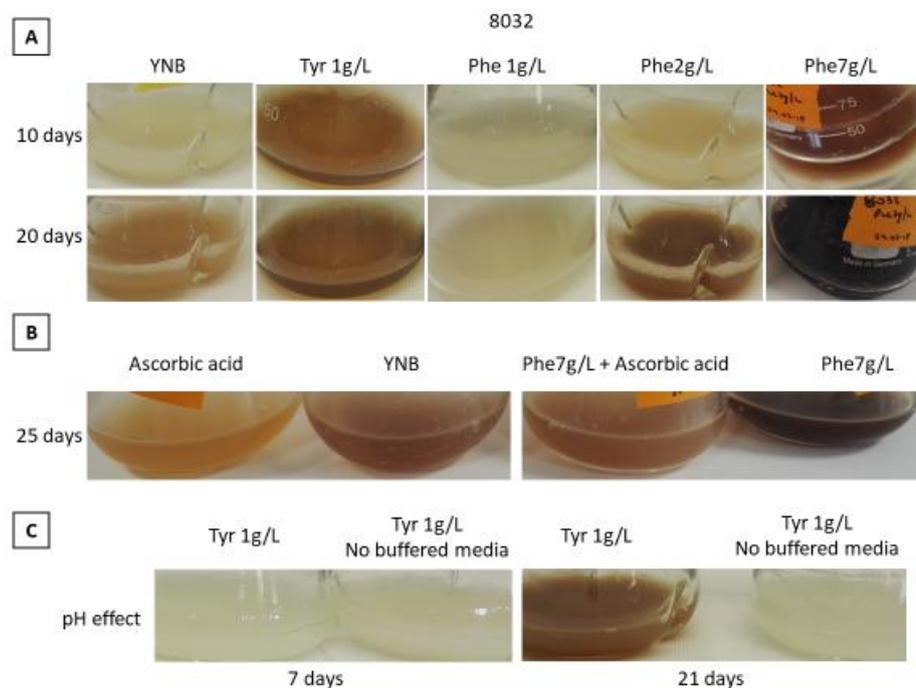
Figure 38. Brown pigment production in different culture conditions

In **A**, it can be observed the evolution of brown colour with time and also depending on the amount and type of precursor added in the culture medium. **B** shows that ascorbic acid has an inhibitory effect on pigment's production. In **C** the inhibitory effect of acid pH, due to the absence of buffer in the media, can be observed. 8032 refers to the strain JMY8032

2013). Therefore, oxidation seems to be an important step on the pigment production as it has been described for pyomelanin synthesis.

Other characteristic of melanin-derived compounds is their insolubility in water, organic solvents and acid solution, and solubility in alkaline solutions (Ye et al., 2011). Thus, in an attempt to extract some pigment from the culture, cells were eliminated after centrifugation and the supernatant was acidified with HCl to pH~2. After 4 hours at room temperature, samples were centrifuged and a brown pellet was obtained which was not soluble on water but soluble in NaOH (Figure 39). These results are also in accordance to what was already described for melanin. This precipitation method allowed us to measure the amount of pigment produced by the strain JMY8032 after 40 days of culture. 1 g/L was quantified doing only one cycle of precipitation.
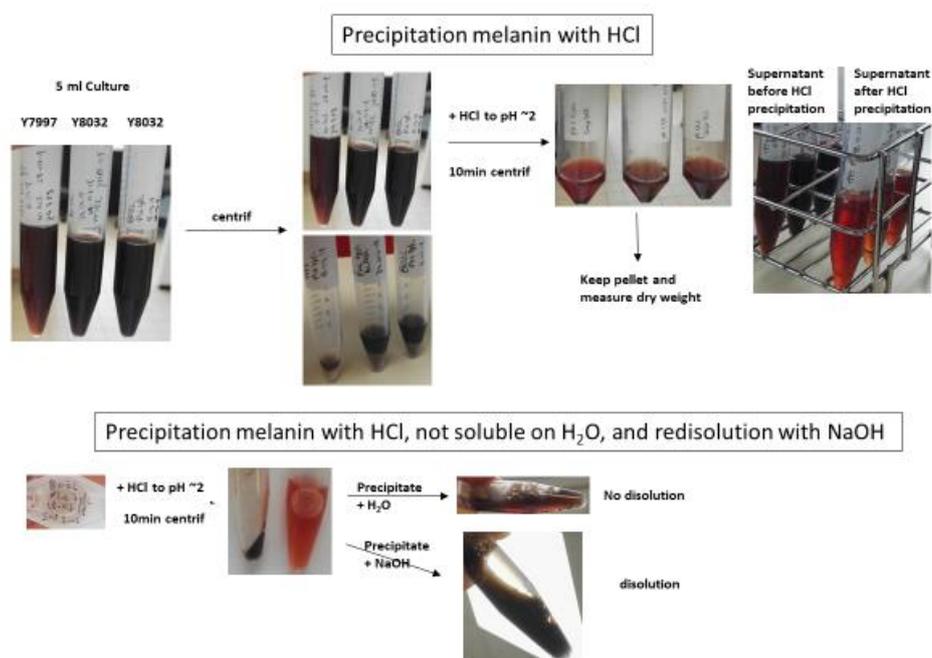
Figure 39. Representation of process for melanin precipitation.

Culture supernatant is acidified and after centrifugation the pyomelanin is recovered as pellet. In the lower part of the figure the solubility of the pellet obtained is tested on H_2O and NaOH.

As previously mentioned, pyomelanin is described to be produced from HGA, which once synthesised in the cell is exported and then autooxidised and polymerised (Figure 37). The gene coding for the enzyme involved in the synthesis of HGA was already annotated in *Y. lipolytica* genome (YALI0B21846g; 4-hydroxyphenylpyruvate dioxygenase, 4HPPD). Therefore, the gene was disrupted or overexpressed in the strain

JMY8032, which produces the brown pigment, aiming to confirm that the pigment produced in the engineered strains is derived from HGA. The gene disruption was achieved by CRISPR-Cas9 technique and, as expected, the obtained strain was not able to produce the brown pigment (Figure 40). On the other hand, the overexpression of the gene increases the brown colour formation (Figure 40). In addition, after the insertion of *4HPPD* cassette, one of the clones appeared to be much browner than the others and colour appears much faster. In order to figure out the cause of this hyperproduction, clones were analysed by RNA-seq. The results showed that all the clones tested had all the expected cassettes, harbouring the genes for overexpressing the AAA pathway and the gene coding for 4HPPD. The hyperproducer strain appears to have a higher expression level of *4HPPD* than the other strains. The DNA sequences permitted to identify the insertion of the cassettes in three different loci of *Y. lipolytica* genome, implying that the hyperproducer strain has three copies of the *4HPPD* gene. This hyperproducing strain can be of great use for the production of high amounts of the pigment.
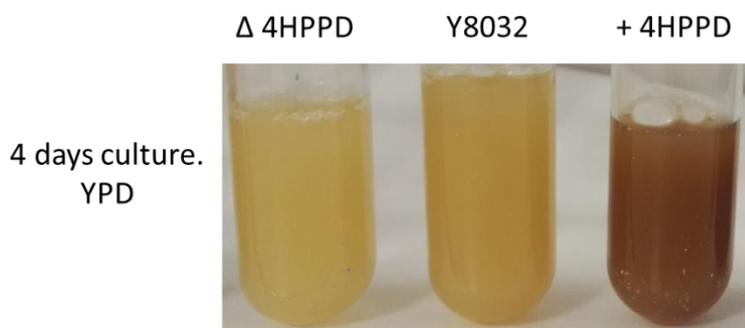
Figure 40. Effect of disruption and overexpression of *4HPPD* in the production of brown pigment
Y8032 = JMY8032: *ARO1-ARO2 -scARO4K229L-scARO7T226I -TKL-ARO8-ARO10.* The deletion of
4HPPD reduce the amount of Brown pigment produced while the overexpression of this gene
increases the pigment in the culture.

Finally, the HGA production was also detected by HPLC in the culture supernatant of engineered strains that produced brown pigments, with higher levels detected on strains engineered to have the higher flux through the shikimate pathway and that produced browner supernatant (Figure 36).

The presented results evidenced that the brown colour produced by the shikimate engineered strains is derived from Tyr and Phe, with HGA as the intermediate involved in the pigment production through oxidation. This strongly suggests that the pigment observed is pyomelanin.

The production of pyomelanin is described in the literature as derived from Tyr (Carreira et al., 2001; Schmaler-Ripcke et al., 2009; Turick et al., 2010), however, I evidenced the production of brown pigment also after the addition of Phe in the medium. As previously mentioned, Phe can be converted irreversibly into Tyr in higher organisms. Even though this was not described in yeasts, some evidence that it might happen in *Y. lipolytica* was previously pointed out in this chapter (end of section 3.2.3.1) after observing the bioconversion of Phe into Tyr-derived Ehrlich metabolites. The fact that the apparition of the brown pigment was slower when the medium was supplemented with Phe than with Tyr may suggest that the production of the pigment occur, in the Phe supplemented medium, after the Phe is transformed into Tyr. Thus, implying that pyomelanin is synthesised from Tyr and that Phe can be converted into Tyr in *Y. lipolytica*.

In addition, by engineering the AAA pathway and the biosynthetic pathway of pyomelanin, a strain with higher and faster production capacities was identified. Culture and extraction conditions should be optimized in order to increase the production, notwithstanding, the biosynthesis of this pigments with a wide variety of properties enlarge the scope of applications of *Y. lipolytica.*

3.2.4 Discussion and conclusion

AAA biosynthetic pathway is a gateway to several useful and interesting metabolites, with a wide range of applications. Thus, the capacity to engineer *Y. lipolytica* for the production of AAA-derived products further diversifies the applications of this organism.

During the present work, several modifications were done in the AAA pathway with the aim of constructing a strain with an increased pool of AAA to be used as a chassis strain for the synthesis of AAA-derived products.

In order to evaluate the constructed strains, a reliable and efficient method capable of measuring the produced compounds was required. Therefore, during this work, three approaches were tested to detect changes on the AAA pathway. The first approach was a promoter-based biosensor system. It was constructed using the promoter region of *ARO10* and a fluorescent protein. Although, the expression of Aro10 was reported to increase in the presence of Phe (Celińska et al., 2015b), in the tested conditions here,

its promoter did not produce any detectable signal. However, further studies can be conducted in order to better understand the regulation of *ARO10*, and to decipher the architecture of the promoter region in order to identify regulatory sequences that can be afterwards engineered to increase its sensitivity and its response strength. In addition, the new identified regulation sequences in the promoter region of other genes of the AAA pathway (Trebulle, 2019), introduced in section 1.4.3, should be studied to know their functionalities and eventually be used as biosensors of the AAA pathway.

The second sensor approach, based on the synthesis of green PVA from Trp, permitted to evidence modifications on the flux through the AAA pathway (Figure 19). However, this approach has some drawbacks as a screening method for this project. First, the precise quantification of the final compound (PVA) in microplate test was not easy nor efficient, which make the technique not reliable when seeking to evaluate small differences of production and limits its application as high-throughput method. Second, due the lack of stable episomal vector for *Y. lipolytica*, the reporter pathway has to be inserted into the genome of this yeast, which has the inconvenient of being randomly inserted into the genome and of not being easy to remove after the screening test. Finally, the pigment is produced from Trp and the aim of this work is to have strains with improved Phe and Tyr, so measuring PVA was not appropriate to evaluate strains with an increased flux after the chorismate branch point. Nevertheless, using this method allow to identify one strain that produces higher amounts of Trp, which can be use as platform for the synthesis of Trp derived products of interest such as violacein and serotonin.

These two tests highlighted the need of stable episomal vectors for *Y. lipolytica*. The availability of such tool will be helpful for the development of biosensor approaches on this yeast given the fact that these systems need to be eliminated of the cells after the screening. The CRISPR vector, which showed to be efficient for *Y. lipolytica*, can be tested as backbone to develop new episomal vectors. Compared to the described episomal vectors used for *Y. lipolytica*, the CRISPR vector only has two small portions (227 bp and 114 bp) of the ARS68 (2310 bp), used for instance in the plasmid used to express the Cre enzyme, and are in separates part of the vector. Nevertheless, a test should be done in order to verify its stability and segregation in the yeast after several growth cycles in order to evaluate its usefulness as stable replicative gene expression system. The availability of reliable, fast, robust and easy to use biosensor are

very useful for high-throughput screening of strains. Thus, efforts on their development should be pursued.

Finally, a HPLC method was set up to detect Ehrlich metabolites derived from Phe and Tyr. However, *ARO8* and *ARO10* have to be overexpressed in order to have a detectable level of these metabolites in the culture supernatant. This third approach was used to evaluate the constructed strains during this work. The same method was also used to detect naringenin and resveratrol.

With the detection method setup, a series of strains were constructed harbouring different modifications on the AAA pathway (Figure 20, Figure 21 and Figure 22). During this work it was observed that the entire AAA pathway need to be overexpressed in order to have a significant increase of the AAA pool. Additionally, the expression of feedback insensitive forms of *S. cerevisiae* enzymes enhanced the flux through the pathway and appears to be more effective than the *Y. lipolytica* corresponding enzymes. In an attempt to increase the available amount of the precursor E4P, *TKL1* was overexpressed. However, its expression in the strain overexpressing the AAA pathway led to a decrease, or very little increase, of the Ehrlich metabolites. This effect can be correlated with the capacity of Tkl1 to catalyse the opposite reaction (Curran et al., 2013) (Figure 41). In *S. cerevisiae* the overexpression of *TKL1* was described to enhance the flux to shikimate pathway (Brückner et al., 2018; Hassing et al., 2019), however this effect was not consistently observed (Liu et al., 2019; Suástegui et al., 2017).

The strain producing the higher amount of Ehrlich metabolites was considered as the best producer of Tyr and Phe (Figure 21). Thus, the equivalent strain without overexpressing *ARO8* and *ARO10* has been built as the first AAA *Y. lipolytica* chassis strain. This strain, JMY8175, overexpresses the endogenous genes *ARO1* and *ARO2* and the heterologous deregulated genes *scARO4^{K229L}*, *sc*ARO3^{K222L} and *scARO7^{T226I}*.

This first AAA chassis strain still can be improved, as well as the culture conditions, to increase the production. With that aim, several points were identified and can be included in a second round of the DBTL cycle.

In this sense, the availability of precursors of the pathways, E4P and PEP, was described as one key limiting factor for the synthesis of shikimate derived products in *S. cerevisiae*, because they are neither abundant nor present in a balanced ratio. Using

^{13}C-metabolic flux analysis (^{13}C-MFA), it was shown in *S. cerevisiae* that the carbon flux towards E4P is at least 1 order magnitude less than towards PEP, indicating that E4P is the rate-limiting starter unit at this stage (Suástegui et al., 2016). In *Y. lipolytica*, PPP and glycolysis are described to be highly expressed (Christen and Sauer, 2011; Wasylenko et al., 2015). However, they should be rewired to increase their entrance into the shikimate pathway (Figure 41). Even though the first attempt to increase the availability of E4P was not successful, other strategies can be tested. For instance, the implementation of a heterologous phosphoketolase pathway, able to split fructo-6-phosphate into E4P and acetyl-phosphate. This introduction could divert part of the carbon flux from the glycolysis toward E4P (Bergman et al., 2016; Hassing et al., 2019; Liu et al., 2019). However, the accumulation of acetate that occurs concomitantly should be prevented, for instance by channelling it towards Acetyl-CoA, in order to avoid a reduction on cell growth (Bergman et al., 2019). The overexpression of *TAL1* and *RKI*, involved in the PPP (Figure 41), can also be evaluated to increase E4P (Mao et al., 2017; Suástegui et al., 2017). In addition, as discuss later, *TAL1* was proposed by an *in-silico* modelling procedure. Approaches such as deletion of *ZWF*, which catalyses the entrance of glucose-6-phosphate into the oxidative PPP (Figure 41), and downregulation of *PYK1*, which converts PEP to pyruvate (Figure 41), can lead to improvements on the flux through the shikimate pathway, however, they were also related to an excessive reduction of growth rate in *S. cerevisiae* (Hassing et al., 2019), because of their key cellular functions. The deletion of *ZWF* block the oxidative PPP, which is the main source of NADPH. On the other hand, downregulation of *PYK* reduced the amount of pyruvate, which is the entrance, through acetyl-CoA, into the TCA cycle (Lehninger et al., 2000).

Regarding the engineering of the AAA pathway, the expression of the AroL enzyme from *E. coli* can be considered. This enzyme catalyses the reaction from shikimate to shikimate-3-phosphate, which is the bottleneck of the five-steps reaction from DAHP to EPSP (Figure 22) catalysed by the pentafunctional enzyme Aro1 in yeasts (Rodriguez

et al., 2015). Its expression showed to have a positive effect on the flux through the AAA pathway in *S. cerevisiae* (Hassing et al., 2019; Rodriguez et al., 2015).

Adapted from Hasunuma *et al.* 2011

Figure 41. Schematic representation of endogenous metabolic pathways connecting glycolysis, PPP, xylose and glycerol utilization as carbon sources and shikimate pathway.

The precursors of the shikimate pathway are circled in red. Bleu circles show possible carbon sources. Green lines highlight enzymes that have been cited in the text as possible targets to increase E4P and PEP. The orange arrow represents the heterologous phosphoketolase pathway. Discontinued arrow represents a multistep pathway. Abbreviations: BPGA, 1,3-bisphosphoglycerate; DHAP, dihydroxyacetonephosphate; E4P, erythrose-4-phosphate; FBP, fructose-1,6-bisphosphate; F6P, fructose-6-phosphate; GAP, glyceraldehyde-3-phosphate; Glycerol3P, glycerol-3-phosphate; G6P, glucose-6-phosphate; PEP, phosphoenolpyruvate; PPP, pentose phosphate pathway; TCA: tricarboxylic acid; 6PG, 6-phosphogluconate; 2PGA, 2-phosphoglycerate; 3PGA, 3-phosphoglycerate; R5P, ribose-5-phosphate; Ru5P, ribulose-5-phosphate; S7P, sedoheptulose-7-phosphate; X5P, xylulose-5-phosphate. Enzymes are indicated by their gene assignment symbols: EYK1: Erythritol kinase; GND, 6-phosphogluconate dehydrogenase; PYK1, pyruvate kinase; SOL, 6-phosphogluconolactonase; TAL, transaldolase; TDH, glyceraldehydes-3-phosphate dehydrogenase; TKL, transketolase; RPE, ribulose-5-phosphate 4-epimerase; RKI, ribose-5-phosphate isomerase; XK, xylulokinase; XDH, xylitol dehydrogenase; XR, xylose reductase; Xfpk, phosphoketolase; ZWF, glucose-6-phosphate dehydrogenase.

As part of the CHASSY project, one of the partners (Chalmers University, Sweden. With Verena Siewers as Project Leader at Jens Nielsen's group) was in charge of developing an enzymatic-constrained genome-scale metabolic model (ecGEMs) (Sánchez et al., 2017) for *Y. lipolytica*. The particularity of these metabolic models is that they take into account enzymes as part of reactions, thereby ensuring that each metabolic

flux does not exceed its maximum capacity. Thus, more accurate predictions of the effect of culture conditions and/or of genomic modifications on the metabolism of *Y. lipolytica* will be possible. When testing this new model called yl-ecGEM, a simulation was done to identify the potential genomic modifications to be done in *Y. lipolytica* in order to increase the pool of AAAs. The results were in accordance with the modifications that were done for the construction of the chassis strain, highlighting the overexpression of *ARO1*, *ARO2*, *ARO4* and *ARO7* as top priority modifications. The model also predicted the overexpression of *GLT1* (glutamate synthase, YALI0B19998g), involved in the synthesis of amino acids, in the nitrogen metabolism and in the energy metabolism, *TAL1* (transaldolase, YALI0F15587g) and *TKL1* (YALI0E06479g), both involved in the PPP pathway and synthesis of E4P (Figure 41). As showed and discussed before, the overexpression of *TKL* in my constructions did not improve the product formation, or very little, however, the co-expression with *TAL1* may have a cooperative effect and force the PPP flux to the production of E4P and consequently increase the AAAs pool. Additionally, the deletion of acyl-transferases, involved in lipids metabolism, and genes involved in serine synthesis pathway, which are involved in amino acid synthesis and redox balance maintenance of the cell, were also proposed by the model. These modifications are now on the pipeline of next engineering strategies to be considered for the construction of the next chassis generation.

Furthermore, culture conditions can be modified in order to increase the pool of AAAs. In *Y. lipolytica* it was described that the use of glycerol as carbon source increases the production of erythritol, whose precursor is the E4P (Carly et al., 2017a). Thus, it can be hypothesised that the disruption of *EYK*, involved in the erythritol biosynthesis, and the culture of the strain in glycerol may increase the availability of E4P for the shikimate pathway (Figure 41). The use of glycerol as carbon source proved to increase the production of 2PE and naringenin in the present work, thus, encouraging the deletion of *EYK*. Glycerol is a very attractive substrate because it is a major byproduct of industrial processes such as the production of biodiesel, and it does not compete with food production as it is the case for glucose.

As well, the use of xylose as carbon source can be investigated since xylose consuming species have more active PPP, thus, are more suitable for shikimate production (M. Gao et al., 2017) (Figure 41). *Y. lipolytica* has the genes for xylose consumption,

however they are not expressed highly enough and extra copies of these genes have to be added into the strain in order to be able to use this carbon source (Niehus et al., 2018).

Overall, combining strain engineering and culture condition optimization can lead to a highly performant chassis strain.

The proper understanding of the metabolism regulation is critical for efficient metabolic engineering since it is the natural way of controlling flux distributions. During the present study, the better understanding of the regulation of the AAA pathway in *Y. lipolytica* was sought by analysing the feedback inhibition of Aro4 and Aro7. On one side, Aro4 appear to be feedback inhibited by Trp and Tyr, and this regulation was alleviated by a point mutation (Aro4^{K221L}). The same Aro4 feedback insensitive variant has been reported recently showing an increase in the titer of the final product in *Y. lipolytica* (Palmer et al., 2020), and it is also in agreement with what is observed in *S. cerevisiae* (Luttik et al., 2008). On the other hand, Aro7 does not seem to be feedback inhibited by Trp or Tyr, and its activity does not change between the wild-type and the mutated form tested here. This implies that regulation behaves differently from *S. cerevisiae* for this enzyme. Nonetheless, adding *ARO7^{G141S}* in a *S. cerevisiae* strain expressing of *ARO4^{K229L}* showed only small differences in the formation of aromatic compounds, suggesting a lower Ki of Aro7 than Aro4 for Tyr (Luttik et al., 2008). No report on the regulation of this enzyme in *Y. lipolytica* has been done to my knowledge until now. The results presented here show that the regulation of this pathway is at least partially different to what has been described in *S. cerevisiae*. This implies that knowledge acquired in *S. cerevisiae* cannot be systematically transposed to *Y. lipolytica*.

It was also observed that the co-expression of Aro4 and Aro7, in their wild-type and mutated forms have a negative effect on the biosynthesis of Ehrlich metabolites (Figure 26 and Figure 27). Additionally, strains harbouring the *ARO4^{K221L}-ARO7^{G139S}* or *ARO4-ARO7* cassettes (JMY8000 and JMY8306) grow around three times slower than the other engineered strains. These results can be related to the emergence of some degree of Trp auxotrophy due to the overexpression of *ARO7*, which decreases the chorismate available for the biosynthesis of this amino acid. This effect was already observed in *S. cerevisiae* (Braus, 1991; Krappmann et al., 2000; Luttik et al., 2008). This *ARO4-ARO7* co-expression negative effect needs to be further studied. In addition,

the regulation of Aro3 in *Y. lipolytica* should also be analysed since the feedback insensitive form of *S. cerevisiae* is more efficient that its isoform, Aro4, and this enzyme is one of the most effective in term of AAA improvement when overexpressed.

Finally, the constructed chassis strain was used for the *de novo* biosynthesis of value-added metabolites: 2PE, naringenin, resveratrol and melanin.

The production of 2PE was increased more than two times when using the chassis strain compared to a strain with native shikimate pathway. This evidenced that not only the last steps of a pathway are important to engineered, in this case overexpression of *ARO8-ARO10*, but also upstream steps to increase the availability of precursors. In addition, it was shown that changing glucose by glycerol as carbon source increases almost twice the amount of this higher alcohol, showing the impact of culture conditions on the biosynthesis of a compound.

The highest amount of *de novo* 2PE production reached during this work was of 0.18 g/L (0.02 mg/OD) using glycerol as carbon source. In this condition, 80% of the total Ehrlich metabolites produced is derived from Phe (Figure 30). Considering that the balance between the alcohol and acid form depends on the redox status of the cell (Figure 6), one could expect that setting appropriate and controlled culture conditions help tilt the PAA produced to 2PE. In addition, the overexpression of the dehydrogenase 2PEdh (YALI0F2493g), which expression is increased in the presence of Phe (Celińska et al., 2015b), can be envisaged in order to favour the reduction of the phenyl acetaldehyde, thus increasing the production of 2PE (Figure 6).

In the same perspective, engineering the prephenate branchpoint to favour the production of Phe over Tyr can be envisaged. This can be done by the overexpressing *PHA2* and/or replacing *TYR1* promoter with a very weak promoter with the aim of directing most of the prephenate to the Phe synthesis while avoiding complete auxotrophy of Tyr (Figure 6).

Interestingly, Celinska et al. showed that the deletion of *BAT2* (aminotransferase) increased by 50% the bioconversion of Phe to 2PE, and suggested a competition between the AAA and BCAA (branched chain amino acids, Leu, Val, Ile) for the NH_3 group of the amino acid, which, in terms of cell physiology is part of a more efficient dissipation of amino groups into different pathways (i.e., Bat2 competes with Aro8 for the source of nitrogen) (Celińska et al., 2019). Therefore, this deletion can be envisaged

for the next round of modifications with the aim of improving the production of 2PE by avoiding losing Phe in a competing pathway.

In order to analyse if the limit of 2PE production was due to a low flux or to other factors such as toxicity of the product, AAAs were added into the medium. Supplementing the medium with Phe or Tyr permitted the increase of the 2PE produced as well as the other Ehrlich metabolites. This evidence that the production of 2PE by the chassis strain is not limited by a saturation of the Ehrlich pathway but limited by the upstream flux. Thus, the chassis strain can still be improved in order to increase the capacity production of 2PE. In addition, the amount of 2PE produced here were below the 2 g/L of 2PE achieved by Celinska and co-workers (Celińska et al., 2013) without toxic effect, thus, almost rejecting the hypothesis of a toxic effect in our production conditions. In addition, it was observed that the bioconversion was more efficient in the wild-type strains than in the AAA engineered strains. This limited bioconversion may be a consequence of regulation mechanisms and need to be further studied in order to understand the mechanisms that take place. This highlight the importance of a better knowledge of the regulatory mechanisms involved in the pathways.

As mentioned, previous works described the production of 2PE by *Y. lipolytica* (Celińska et al., 2019, 2015b, 2013). Celinska and co-workers obtained the highest titer of 2PE described for *Y. lipolytica*, using a wild-type strain. They reached 2 g/L of 2PE by adding 7 g/L of Phe during the stationary growth phase. However, when Phe was added at the inoculation point of culture, they obtained 0.77 g/L of 2PE, evidencing that the Phe added to the media is being used by other competing pathways, e.g., for growth (Celińska et al., 2013). Additionally, in a culture using 50 g/L glycerol as carbon source and supplemented with 2 g/L Phe, they reached 0.42 g/L of 2PE (0.04 mg/OD) in a strain that overexpressed *ARO8*, *ARO10* and a dehydrogenase to favor the ethanol form. The production was increased to 0.6 g/L (0.08mg/OD) when *BAT2*, which compete for Phe catabolism, was deleted (Celińska et al., 2019).

Even though these previous works showed two to ten times higher total production of 2PE than in the present work, it is important to notice that in these works the production was done by bioconversion, meaning that Phe was added in the media in order to convert it into 2PE, which has the inconvenient of making the process not cost effective.

While in the present work, the *de novo* synthesis was studied, without adding any supplement and using the glycerol as low-cost carbon source.

Regarding naringenin and resveratrol, their biosynthesis was achieved on strains with engineered shikimate pathway, proving that the heterologous pathway was functional in *Y. lipolytica*.

However, the low amounts obtained and gene expression results revealed some problems on the constructed strains. On one hand, the chassis strain lost some expression cassettes after adding the Cre recombinase enzyme to recover selection markers. On the other hand, the expression level of most of the heterologous genes from the naringenin and resveratrol pathways tested are relatively low. As a result, a new run of production strains should be constructed taking the previous results into consideration. The heterologous pathway should be constructed again in the AAA chassis strain, and the level of expression of heterologous genes from the naringenin and resveratrol pathways must be increased. As previously mentioned, in addition to promoter strength, the choice of orthologs genes play a tremendous role in the final titers. Thus, in order to accelerate the identification of the best combination of genes and promoters for the expression of the heterologous pathway, a combinatorial test can be conducted taking advantage of the GG assembly technique. To identify high producers, appropriate sensors with high sensitivity are necessary for screening the resulting libraries, thus, the previously described naringenin biosensor (Trabelsi et al., 2018) can be very helpful to identify naringenin producer strains.

In addition, further engineering can be envisaged, such as deletion of *ARO10* to avoid the catabolism of Phe and Tyr by the Ehrlich metabolism, or reduction of lipid biosynthesis which may increase the availability of malonyl-CoA for the biosynthesis of flavonoids.

Finally, the chassis strain resulted in good *de novo* production of pyomelanin. This brown pigment was characterized and identified by physicochemical and genetic tests. In addition, by engineering the biosynthetic pathway of pyomelanin, a strain with higher and faster production capacities was identified. Culture and extraction conditions should be optimized in order to increase the productivity, notwithstanding, the biosynthesis of this AAA-derived pigment with a wide variety of properties enlarge the scope of potential applications of *Y. lipolytica.*

Regarding the production of pyomelanin in *Y. lipolytica*, Carreira et al. measured the brown colour at 400nm and achieved a maximal absorbance of 2.5 in a media at pH 7.4 and with 1.5 mM of the precursor HGA (Carreira et al., 2001b). Ben Tahar and co-workers obtained a maximum yield of 0.57 g/L (corresponding to theoretical OD_{400} of 5.8) of pyomelanin using a medium supplemented with Tyr. In the present work, after a non-optimized culture conditions and non-optimized extraction method, the amount of *de novo* pyomelanin produced was of around 1 g/L using the strain Y8032 after 40 days of culture on minimal YNB. This amount can be easily increased using more per-formant strains, e.g. the hyperproducer detected clone with three copies of 4*HPPD*, optimizing culture parameters such as pH, oxygen and temperature that have been found to contribute to productivity (Carreira et al., 2001b; Martínez et al., 2019), and optimising extraction, e.g. doing several rounds of culture supernatant precipitation. This highlight the capacity of *Y. lipolytica* as a melanin producer organism.

As mentioned before, the AAA pathway was relatively unexplored in *Y. lipolytica* up to now. In addition to the already mentioned works of Celinska and co-workers (Celińska et al., 2019, 2015b, 2013) in which differentially expressed proteins in the presence of Phe were identified and later the function of *ARO8* and *ARO10* were deter-mined, recent works did minor modification in the AAA pathway for increasing the production of heterologous aromatic derived compounds. Lv and co-workers increased the production of naringenin by 50% by overexpressing *ARO1* (Lv et al., 2019b), Palmer and co-workers expressed a feedback-insensitive form of Aro4 and saw a four-fold increase on their naringenin production (Palmer et al., 2020) and Shang and co-workers increased eight times the amount of arbutin produced by overexpressing *ARO3* and *ARO4* (Shang et al., 2020). However, these works did not seek to exploit the com-plete engineering of the AAA pathway, nor to construct a platform chassis strain for the production of a variety of product derived from AAA. In addition, it is to notice that all these referenced works producing heterologous aromatic compounds cultured their tested strains in rich complete medium, YPD, or added a complete supplement mixture to the culture, which contain Phe and Tyr in their formulations. The presence of these amino acids in the media favours the synthesis of compounds derived from them. The strains tested during this thesis, on the contrary, were grown in minimal medium with-out any supplementation.

Further improvements on the chassis strain can be done in order to reach higher levels of the pool of AAA, thus, increasing its efficiency as cell factory. Also, more studies are needed to better understand the regulation of the AAA pathway in *Y. lipolytica*. However, the chassis strain constructed in this work proved to have a higher flux through the AAA pathway compared to the wild-type strain and its usefulness for the *de novo* biosynthesis of added-value molecules.

4 General conclusion

The use of microorganisms as cell factories for the production of molecules of interest is appearing as an attractive alternative to chemical synthesis, whose harsh conditions, wastes and oil-derived precursors make of it an unsustainable process. In this regard, bio-sourced molecules can provide an environmentally friendly process by using low-cost, non-food feedstock to produce molecules of interest. With the advent of inexpensive DNA sequencing and synthesis and bioinformatic tools this solution is feasible, albeit moving cell factories from a proof-of-concept to production strains that meet the industrial requirements is challenging and still requires multiple rounds of the DBTL cycle. Additionally, the expansion of strains used as cell factories, toolbox available for genome engineering, high-throughput screening methods and the range of molecules obtained is essential in order to find robust and cost-effective processes.

With the aim of contributing to increase the availability of tools and strains that can be used as cell factories, this work proposes the non-conventional yeast *Y. lipolytica* as a framework for the construction of chassis strains for the production of AAA-derived compounds, which have a plethora of applications regarding human health and nutrition. Non-conventional yeasts can provide many potential advantages over *S. cerevisiae* with respect to substrates, pathway requirements, product profile, and gross physiology. Therefore, their fundamental knowledge and their use as cell factory need to be expanded. In order to improve our capacity to modify the metabolism of this yeast, synthetic biology tools based on standardized, characterized and reusable components were also developed during this work.

Thus, the identification of a problem or need, and setting the objectives and strategies to solve it, can be the starting point, the *design*, of a DBTL cycle. The approaches carried out during the present work to achieve the objectives are part of the *build* and *test* parts of the cycle. The obtained results will finally give insight on what should be done in order to improve the system, these perspectives for future work can be identified as part of *learn* part of the cycle (Figure 42).
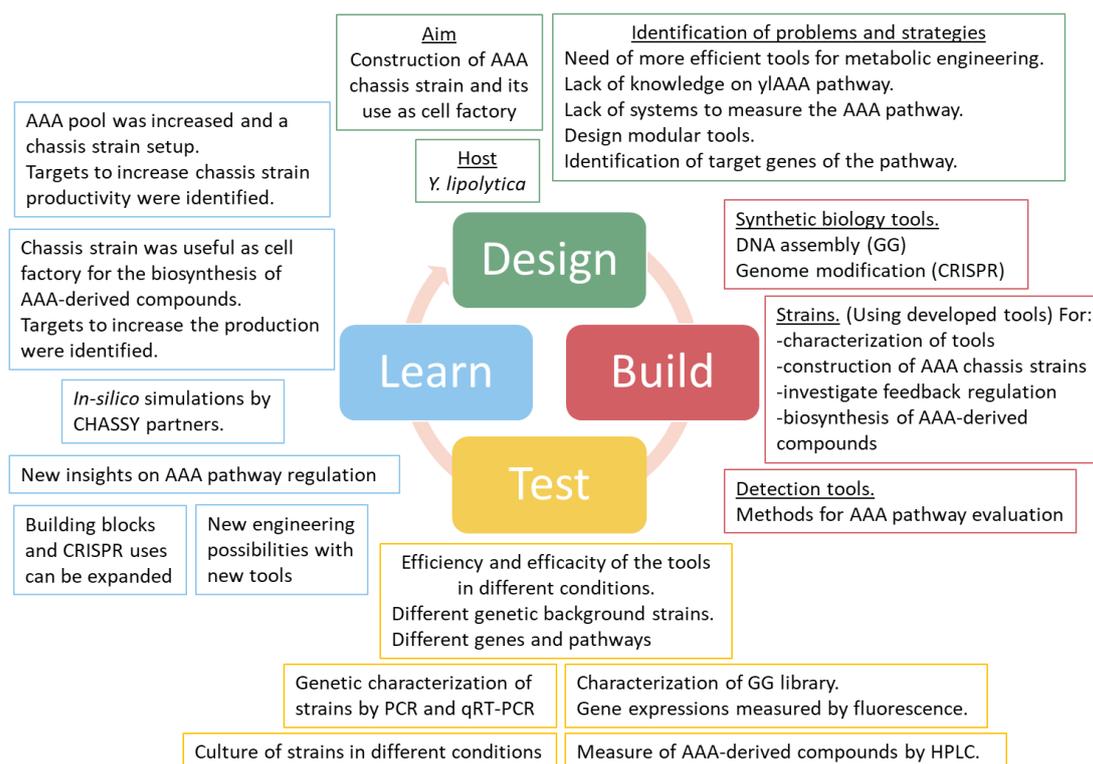
Figure 42. The DBTL cycle during this work

The different strategies of the presented work are part of a DBTL cycle. The acquired knowledge will feed the next round of the cycle with the aim of improving the efficiency of the cell factories and the basic knowledge on *Y. lipolytica* physiology.

As stated earlier, the ability to engineer the metabolism of a microorganism depend on the availability of molecular biology tools. This is one of the bottlenecks when using non-conventional organisms. Thus, with the objective of increasing the efficiency when engineering *Y. lipolytica* metabolism, tools that can be applied in a generic systematic manner to construct expression cassettes and modify the genome of *Y. lipolytica* were developed during this work and presented in chapter 3.

For DNA assembly, a modular cloning technique based on GG was set-up. This method allows to assemble in a one pot reaction multiple and interchangeable DNA fragments by applying digestion-ligation cycles. Thanks to IIS restriction enzymes, a scaffold of predesigned 4 nt was designed to allow the assembly of 13 modular genetic elements, including: 1, 2 or 3 TUs (each consisting of promoter, gene and terminator), a selective marker and sequences for genome integration. Furthermore, the system contains a color-based reporter gene (RFP) to reduce the screening of false positive clones after the assembly. To expand the usefulness of the system, a toolkit composed of 64 interchangeable building blocks was constructed and characterized. The constructed library covers the 13 modular genetic elements and allows the construction of expression

cassettes with different levels of gene expression, that can be integrated in different loci of the genome and in several genetic background strains. In addition, the GG system proved to be very efficient for the rapid construction of strains with complex metabolic pathways by constructing carotenoids producing strains and strains capable of using xylose as carbon source. Furthermore, the combinatorial capacity of this technique was used for shuffling promoters on the three-gene carotenoid heterologous pathway, allowing the fast identification of best promoter-gene combination for each position that ensure the best production capacity of the pathway.

The several advantages of this GG method highly increase the possibilities regarding metabolic engineering of *Y. lipolytica*. The reusable library of characterized DNA fragments, together with the capacity of limitless expansion of the library and the efficient shuffling of parts allowed by the combinatorial approach make this method very efficient for the fast assembly of large variety of bio-bricks. It has been established to consolidate a long-term strategy of synthetic biology in *Y. lipolytica*. The huge combinations of parts that can be rapidly constructed and explored is a great advantage for the identification of best producer strain in a cell factory project and also for more fundamental research in order to identify or characterise traits of this yeast. Furthermore, the capacity of the system to assemble multiple genes and to allow the design of cassettes to be inserted at a specific site in the genome, allowing the simultaneous deletion of a competing pathway, has the advantage of reducing the rounds of transformations needed to reach a desired construction, which increases the viability and/or transformability of recombinant strains.

For genome editing, a versatile and modular CRISPR-Cas system was set up. The vector containing the system is constructed based on the GG method, thus, allowing the easy change of markers, Cas promoter and gRNA, and the recycling of GG bricks. A set of six CRISPR-Cas9 platform vectors, each containing different selection makers, was constructed to easily introduce any gRNA by GG assembly. Here also, the RFP is used as reporter gene for reducing screening of false positive. The developed toolkit can be used in a wide range of different *Y. lipolytica* genetic background including wild-type strains. The use of CRISPR-Cas9 in *Y. lipolytica* improved the HR from 20% to 80% and permitted to reduce the length of the homologous recombination region from 1000 to 100 bp. The vector can also be used to knock out genes with the objective of

increasing the panel of markers for this yeast. Additionally, based on the same vector design, a CRISPR-Cpf1 system was constructed and proved to be functional for the disruption of *Y. lipolytica* genome.

Even though efforts were done, the markerless insertions as well as the multiplex strategies should be improved. Reducing the number of markers needed and increasing the number of simultaneous modifications that can be done, will reduce time and increase the possibilities of genome engineering. Furthermore, other Cas enzymes exists, with different features, and their development can help to increase the scope of applications of the CRISPR systems. In addition, efforts are still needed in order to better understand the efficiency of gRNA whose predictability is not optimal yet.

The fast and combinatorial assembly of complex synthetic pathway together with efficient and site-specific genome editing capability greatly enriches the spectrum of possible studies that can be conducted in *Y. lipolytica*. Additionally, the site-directed modifications that can be achieved with these techniques allows the construction of better characterised strains, which is required for industrial yeasts.

Both toolkits are available to the community through Addgene and the detailed protocols published in the associated scientific articles. Furthermore, the standardized design of building blocks that form both systems allows the easy expansion of them and facilitates the collaboration with other teams. These new developed tools enabled the fast construction of strains developed on the second part of this work, described along chapter 4.

With the purpose of developing a *Y. lipolytica* chassis strain with an increased pool of AAA, particularly of Phe and Tyr, a large series of strains was constructed.

In order to detect changes on the AAA pathway of the engineered strains, three approaches were tested during this work. However, only the HPLC method proved to be useful for the precise evaluation of the constructed strain. On the other side, the *ARO10* promoter-based biosensor did not produce any detectable signal in the conditions tested, and the use of PVA, derived from Tyr, as colour reporter molecule, was not efficient for the detection of modifications on the pool of Tyr and Phe.

Even though HPLC is highly precise, it is not particularly suited for high-throughput screening of modified strains. Thus, efforts should be pursued in order to have a

method that permit the rapid and easy evaluation of the constructed strain. This will allow to explore large strains libraries, for instance obtained as a result of GG-parts shuffling, and easily identify improved phenotypes, in the present case, the best Phe/Tyr producer strain.

For the construction of engineered strains, genes involved in the AAA pathway were adapted for GG and expressed in *Y. lipolytica*. The differentially engineered strains were cultured and the Ehrlich metabolites produced and secreted into the medium were measured by HPLC as reporter metabolites of the intracellular AAAs. The results evidenced that the AAA pathway have to be entirely overexpressed in order to have a significant increase of the pool of Phe and Tyr. Additionally, the expression of feedback insensitive enzymes from *S. cerevisiae* (scAro4^{K229L} and scAro7^{T226I}) increase the carbon flow compared to the wild-type forms from *Y. lipolytica*. On the contrary, the overexpression of *TKL1*, intended to increase the precursor E4P, may have preference by the opposite reaction thus reducing the flux through the pathway. These differentially engineered strains permitted to identify a strain producing higher amounts of Phe and Tyr than the others, JMY8175 expressing *ARO1-ARO2-scARO4^{K229L}-scARO3^{K222L}-scARO7^{T226I}*. As discussed in chapter 4, this first AAA chassis strain can be further improved. Several alternatives were presented as possible targets for future metabolic engineering as part of the second round of the DBTL cycle, also possible culture conditions were highlighted. For instance, the shikimate precursors availability can be increased by engineering the PPP, using glycerol as carbon source and inhibiting the competing erythritol pathway; also, the pull of precursors into the shikimate pathway can be favoured by increasing the flux through the pathway using deregulated forms of Aro3 and overexpressing enzymes involved in the limiting steps of the pathway like ecAroL. Efforts should be pursued in this regard in order to increase the role of *Y. lipolytica* as AAA cell factory.

In order to increase the knowledge on the regulation of *Y. lipolytica* AAA pathway, the feedback regulation of *Y. lipolytica* Aro4 and Aro7 was tested. To do so, the enzymes were tested *in vitro* and *in vivo*, and were mutated in order to imitate the deregulated forms of the orthologs enzymes in *S. cerevisiae*. The results showed that Aro4 of *Y. lipolytica* is repressed by Trp and Tyr, and that the point mutation done (K221L) alleviate this regulation, resulting in a more active enzyme, similarly to *S. cerevisiae*

Aro4. On the contrary, Aro7 does not seems to be feedback regulated by Trp or Tyr in *Y. lipolytica*, differing from the activity profile of the *S. cerevisiae* enzyme. During this test, it was also observed that the co-expression of Aro4 and Aro7 from *Y. lipolytica* had a deleterious effect on the pool of AAAs. This could be due to a Trp auxotrophic effect caused by an increased flow of chorismate towards Phe and Tyr, nonetheless, it must be further investigated. The study done in this thesis work is, to my knowledge, the first with the aim to better understand the regulation of the AAA pathway in *Y. lipolytica*.

Finally, the AAA chassis strains were used for the *de novo* biosynthesis of endogenous and heterologous value-added compounds derived from Phe and Tyr: 2PE, a rose-like aroma compound, naringenin, a flavonoid, and melanin, an UV and ROS protector. The *de novo* production of 2PE was possible after adding two extra genes at the chassis strain, *ARO8* and *ARO10*, achieving 0.18 g/L. It was shown that glycerol as carbon source was more efficient than glucose, possibly due to the capacity of glycerol to increase the pool of E4P, a precursor of the shikimate pathway. Even though the amount produced in *Y. lipolytica* is still below the production in *S. cerevisiae*, 2PE production can be increased by modifying the equilibrium among the Ehrlich metabolites, as 2PE is only one of the four metabolites obtained from Ehrlich metabolism of Phe and Tyr, and by further improvements of chassis strain as indicated above.

Regarding naringenin and resveratrol, the produced amounts obtained were very low due to the loss of some of the genes of the engineered AAA pathway and to a relatively low expression of the heterologous genes. However, as the causes were identified, the quick remediation is possible and is undergoing. Despite the default identified in our strains, some naringenin and resveratrol are produced, which is encouraging to pursue with further modifications of the pathway in order to increase the production level. Thus, besides the mentioned repairs, a shuffling of promoters and orthologs genes is already setup with the combination of 16 different genes coming from different organisms.

The engineered chassis strain proved to be a good producer of pyomelanin. The brown pigment was identified and characterized by physicochemical analysis and by overexpression and deletion of a gene involved in its biosynthetic pathway. A hyperproducer strain was identified capable of producing pyomelanin much faster than the other

strains. Culture condition and the extraction methods can be optimized to favour the production of this pigment; nonetheless, 1 g/L was obtained during this work.

Even though very recent works did minor modifications in the AAA pathway of *Y. lipolytica* (Lv et al., 2019b, 2019a; Palmer et al., 2020), none of them aimed to construct a chassis strain with an increased pool of AAA, nor the exploration of the regulation and engineering of this pathway.

As mentioned, a AAA chassis strain was constructed and proved to be useful for the biosynthesis of different molecules. However, the constructed strains can be further improved. In that sense, several options were previously discussed in order to increase their specific molecule production. Increasing the pool of AAA, will increase the value of this *Y. lipolytica* chassis strain, allowing higher production level of final compounds and expanding its use for to the production of other metabolites of interest derived from the AAA pathway. Additionally, further efforts should be done in order to increase the knowledge on the regulation of this pathway. Understanding how metabolism is regulated may allow different strategies for pathway engineering, such as targeting regulation instead of overexpressing specific pathways. In this regard, the regulation of Aro3 should be evaluated, as it is described to be more efficient than the isoenzyme Aro4 in *S. cerevisiae*. Moreover, it may allow to take over the negative effect observed when *Y. lipolytica* Aro4 and Aro7 were co-expressed. Also, in this work it was evidenced a possible conversion of Phe into Tyr. This effect, known for higher organism, was, to my knowledge, not described before in yeasts. The confirmation of this conversion by a dedicated test or larger amount of data should be follow by a characterisation of the mechanisms involved. Knowing, for instance, the enzyme involved will allow to engineer the pathway also at that point to favour one or another AAA.

In addition, further enlarging the toolbox dedicated for this yeast is of great importance. For instance, to increase the engineering possibilities on this yeast, to be able to construct strains that are more robust and better characterized, and to expand the evaluation techniques. To do so, new methods can be developed or adapted from existing ones, also the efficiency and capabilities of the methods available for *Y. lipolytica* can be enhanced. Thus, efforts should be addressed to better understand CRISPR efficiency, to continue the development of Cpf1 and other Cas enzymes system, as well as to increase the number and type of building blocks such as selection markers to limit

iterative marker curation. In addition, the development of stable episomal plasmids, which are useful to rapid and temporary express genes, for instance for biosensor development, is still a debt in *Y. lipolytica* toolbox.

Pursuing efforts in this sense will continue to facilitate studies in this yeast, thus, increasing the knowledge on its physiology, as well as offering new perspectives in metabolic pathway engineering and in potential biotechnological applications. In doing so, *Y. lipolytica* enlarges its possibilities as cell factory and will also soon no longer be considered as a non-conventional yeast.

Overall, this work successfully expanded the toolbox dedicated to *Y. lipolytica* with two robust, fast and versatile techniques, one for DNA assembly and the other for site-specific genome editing. These techniques proved to be highly useful for the rapid construction of optimized cell factories. Furthermore, the optimization of the AAA pathway, not much explored in *Y. lipolytica*, enlarges the scope of possible applications of this industrially relevant microorganism and at the same time increases the knowledge on how this metabolic pathway may function in this yeast. The construction of a AAA chassis strain proved to be useful for the production of different compounds with a variety of applications. The acquired information will feed the DBTL cycle and be used to improve next generation of aromatics cell factories and experiments design, as well as the general metabolic model of *Y. lipolytica*, which in turn could be applied to other studies and applications.

5 Bibliography

Adams, B.L., 2016. The Next Generation of Synthetic Biology Chassis: Moving Synthetic Biology from the Laboratory to the Field. ACS Synth. Biol. 5(12):1328-1330. https://doi.org/10.1021/acssynbio.6b00256

Ahn, S.Y., Choi, M., Jeong, D. woon, Park, S.A., Park, H.A., Jang, K.S., Choi, K.Y., 2019. Synthesis and chemical composition analysis of protocatechualdehyde-based novel melanin dye by 15T FT-ICR: High dyeing performance on soft contact lens. Dye. Pigment. 160, 546–554. https://doi.org/10.1016/j.dyepig.2018.08.058

Apte, M., Girme, G., Bankar, A., Ravikumar, A., Zinjarde, S., 2013a. 3, 4-dihydroxy-L-phenylalanine-derived melanin from *Yarrowia lipolytica* mediates the synthesis of silver and gold nanostructures. J. Nanobiotechnology.11:2 https://doi.org/10.1186/1477-3155-11-2

Apte, M., Sambre, D., Gaikawad, S., Joshi, S., Bankar, A., Kumar, A.R., Zinjarde, S., 2013b. Psychrotrophic yeast *Yarrowia lipolytica* NCYC 789 mediates the synthesis of antimicrobial silver nanoparticles via cell-associated melanin. AMB Express 3, 1–8. https://doi.org/10.1186/2191-0855-3-32

Awan, A.R., Blount, B.A., Bell, D.J., Shaw, W.M., Ho, J.C.H., McKiernan, R.M., Ellis, T., 2017. Biosynthesis of the antibiotic nonribosomal peptide penicillin in baker's yeast. Nat. Commun. 8. https://doi.org/10.1038/ncomms15202

Bae, S.J., Park, B.G., Kim, B.G., Hahn, J.S., 2019. Multiplex Gene Disruption by Targeted Base Editing of *Yarrowia lipolytica* Genome Using Cytidine Deaminase Combined with the CRISPR/Cas9 System. Biotechnol. J. 15(1):e1900238. https://doi.org/10.1002/biot.201900238

Balderas-Hernández, V.E., Sabido-Ramos, A., Silva, P., Cabrera-Valladares, N., Hernández-Chávez, G., Báez-Viveros, J.L., Martínez, A., Bolívar, F., Gosset, G., 2009. Metabolic engineering for improving anthranilate synthesis from glucose in *Escherichia coli*. Microb. Cell Fact. 8, 19. https://doi.org/10.1186/1475-2859-8-19

Bankar, A. V., Kumar, A.R., Zinjarde, S.S., 2009. Environmental and industrial applications of *Yarrowia lipolytica*. Appl. Microbiol. Biotechnol. 84(5):847-865 https://doi.org/10.1007/s00253-009-2156-8

Barth, G., Beckerich, J.-M., Dominguez, A., Kerscher, S., Ogrydziak, D., Titorenko, V., Gaillardin, C., Hohmann, S., 2003. Functional genetics of *Yarrowia lipolytica*, in: de Winde, J.H. (Ed.), Functional Genetics of Industrial Yeasts. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 227–271.

Barth, G., Gaillardin, C., 1997. Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. FEMS Microbiol. Rev. 19, 219–237.

Barth, G., Gaillardin, C., 1996. *Yarrowia lipolytica*, in: Nonconventional Yeasts in Biotechnology. Springer Berlin Heidelberg, pp. 313–388.

Ben-David, Y., Zlotnik, E., Zander, I., Yerushalmi, G., Shoshani, S., Banin, E., 2018. SawR a new regulator controlling pyomelanin synthesis in *Pseudomonas aeruginosa*. Microbiol. Res. 206, 91–98. https://doi.org/10.1016/j.micres.2017.10.004

Ben Tahar, I., Kus-Liśkiewicz, M., Lara, Y., Javaux, E., Fickers, P., 2019. Characterization of a nontoxic pyomelanin pigment produced by the yeast *Yarrowia lipolytica*. Biotechnol. Prog. 36(2), e2912. https://doi.org/10.1002/btpr.2912

Beopoulos, A., Verbeke, J., Bordes, F., Guicherd, M., Bressy, M., Marty, A., Nicaud, J.-M., 2014. Metabolic engineering for ricinoleic acid production in the oleaginous yeast *Yarrowia lipolytica*. Appl. Microbiol. Biotechnol. 98, 251–262. https://doi.org/10.1007/s00253-013-5295-x

Bergman, A., Hellgren, J., Moritz, T., Siewers, V., Nielsen, J., Chen, Y., 2019. Heterologous phosphoketolase expression redirects flux towards acetate, perturbs sugar phosphate pools and increases respiratory demand in Saccharomyces cerevisiae. Microb. Cell Fact. 18, 25. https://doi.org/10.1186/s12934-019-1072-6

Bergman, A., Siewers, V., Nielsen, J., Chen, Y., 2016. Functional expression and evaluation of heterologous phosphoketolases in *Saccharomyces cerevisiae*. AMB Express 6. https://doi.org/10.1186/s13568-016-0290-0

Blazeck, J., Hill, A., Jamoussi, M., Pan, A., Miller, J., Alper, H.S., 2015. Metabolic engineering of *Yarrowia lipolytica* for itaconic acid production. Metab. Eng. 32, 66–73. https://doi.org/10.1016/j.ymben.2015.09.005

Blazeck, J., Liu, L., Redden, H., Alper, H., 2011. Tuning Gene Expression in *Yarrowia lipolytica* by a Hybrid Promoter Approach ▽ . Appl. Environ. Microbiol. 77, 7905–7914. https://doi.org/10.1128/AEM.05763-11

Bordes, F., Fudalej, F., Dossat, V., Nicaud, J.M., Marty, A., 2007. A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. J. Microbiol. Methods 70, 493–502. https://doi.org/10.1016/j.mimet.2007.06.008

Bothma, J.P., De Boor, J., Divakar, U., Schwenn, P.E., Meredith, P., 2008. Device-quality electrically conducting melanin thin films. Adv. Mater. 20, 3539–3542. https://doi.org/10.1002/adma.200703141

Braga, A., Belo, I., 2016. Biotechnological production of γ-decalactone, a peach like aroma, by *Yarrowia lipolytica*. World J. Microbiol. Biotechnol. 32, 169. https://doi.org/10.1007/s11274-016-2116-2

Braus, G.H., 1991. Aromatic amino acid biosynthesis in the yeast *Saccharomyces cerevisiae*: a model system for the regulation of a eukaryotic biosynthetic pathway. Microbiol. Rev. 55, 349–370.

Brophy, J.A.N., Voigt, C.A., 2014. Principles of genetic circuit design. Nat. Methods. https://doi.org/10.1038/nmeth.2926

Bru, R., Sellés, S., Casado-Vela, J., Belchí-Navarro, S., Pedreño, M.A., 2006. Modified cyclodextrins are chemically defined glucan inducers of defense responses in grapevine cell cultures. J. Agric. Food Chem. 54, 66–71. https://doi.org/10.1021/jf051485j

Brückner, C., Oreb, M., Kunze, G., Boles, E., Tripp, J., 2018. An expanded enzyme toolbox for production of cis, cis-muconic acid and other shikimate pathway derivatives in *Saccharomyces cerevisiae*. FEMS Yeast Res. 18. https://doi.org/10.1093/femsyr/foy017

Carly, F., Gamboa-Melendez, H., Vandermies, M., Damblon, C., Nicaud, J.-M., Fickers,

P., 2017a. Identification and characterization of EYK1, a key gene for erythritol catabolism in *Yarrowia lipolytica*. Appl. Microbiol. Biotechnol. 101, 1–10. https://doi.org/10.1007/s00253-017-8361-y

Carly, F., Vandermies, M., Telek, S., Steels, S., Thomas, S., Nicaud, J.-M., Fickers, P., 2017b. Enhancing erythritol productivity in *Yarrowia lipolytica* using metabolic engineering. Metab. Eng. 42, 19–24. https://doi.org/10.1016/j.ymben.2017.05.002

Carreira, A., Ferreira, L.M., Loureiro, V., 2001a. Production of brown tyrosine pigments by the yeast *Yarrowia lipolytica*. J. Appl. Microbiol. 90, 372–9. https://doi.org/10.1046/j.1365-2672.2001.01256.x

Carreira, A., Ferreira, L.M., Loureiro, V., 2001b. Brown Pigments Produced by *Yarrowia lipolytica* Result from Extracellular Accumulation of Homogentisic Acid. Appl. Environ. Microbiol. 67, 3463–3468. https://doi.org/10.1128/AEM.67.8.3463-3468.2001

Carreira, A., Ferreira, L.M., Loureiro, V., 1998. Pigment producing yeasts involved in the brown surface discoloration of ewes' cheese. Int. J. Food Microbiol. 41, 223–230. https://doi.org/10.1016/S0168-1605(98)00054-3

Caruso, D., Berra, B., Giavarini, F., Cortesi, N., Fedeli, E., Galli, G., 1999. Effect of virgin olive oil phenolic compounds on in vitro oxidation of human low density lipoproteins. Nutr. Metab. Cardiovasc. Dis. 9, 102–107.

Celińska, E., Białas, W., Borkowska, M., Grajek, W., 2015a. Cloning, expression, and purification of insect (*Sitophilus oryzae*) alpha-amylase, able to digest granular starch, in Yarrowia lipolytica host. Appl. Microbiol. Biotechnol. 99, 2727–2739. https://doi.org/10.1007/s00253-014-6314-2

Celińska, E., Borkowska, M., Białas, W., Korpys, P., Nicaud, J.-M., 2018. Robust signal peptides for protein secretion in *Yarrowia lipolytica*: identification and characterization of novel secretory tags. Appl. Microbiol. Biotechnol. 102, 5221–5233. https://doi.org/10.1007/s00253-018-8966-9

Celińska, E., Borkowska, M., Białas, W., Kubiak, M., Korpys, P., Archacka, M., Ledesma-Amaro, R., Nicaud, J.-M., 2019. Genetic engineering of Ehrlich pathway modulates production of higher alcohols in engineered *Yarrowia lipolytica*. FEMS

Yeast Res. 19. https://doi.org/10.1093/femsyr/foy122

Celińska, E., Grajek, W., 2013. A novel multigene expression construct for modification of glycerol metabolism in *Yarrowia lipolytica*. Microb. Cell Fact. 12, 102. https://doi.org/10.1186/1475-2859-12-102

Celińska, E., Kubiak, P., Białas, W., Dziadas, M., Grajek, W., 2013. *Yarrowia lipolytica*: The novel and promising 2-phenylethanol producer. J. Ind. Microbiol. Biotechnol. 40, 389–392. https://doi.org/10.1007/s10295-013-1240-3

Celińska, E., Ledesma-Amaro, R., Larroude, M., Rossignol, T., Pauthenier, C., Nicaud, J.-M., 2017. Golden Gate Assembly system dedicated to complex pathway manipulation in *Yarrowia lipolytica*. Microb. Biotechnol. 10(2):450-455.. https://doi.org/10.1111/1751-7915.12605

Celińska, E., Olkowicz, M., Grajek, W., 2015b. L-Phenylalanine catabolism and 2-phenylethano synthesis in *Yarrowia lipolytica*-mapping molecular identities through whole-proteome quantitative mass spectrometry analysis. FEMS Yeast Res. 15. https://doi.org/10.1093/femsyr/fov041

Cheng, F., Tang, X.L., Kardashliev, T., 2018. Transcription Factor-Based Biosensors in High-Throughput Screening: Advances and Applications. Biotechnol. J. 13(7):e1700648. https://doi.org/10.1002/biot.201700648

Chi, H., Wang, X., Shao, Y., Qin, Y., Deng, Z., Wang, L., Chen, S., 2019. Engineering and modification of microbial chassis for systems and synthetic biology. Synth. Syst. Biotechnol. 4, 25-33 https://doi.org/10.1016/j.synbio.2018.12.001

Choi, S.Y., Yoon, K.H., Lee, J. Il, Mitchell, R.J., 2015. Violacein: Properties and production of a versatile bacterial pigment. Biomed Res. Int. https://doi.org/10.1155/2015/465056

Christen, S., Sauer, U., 2011. Intracellular characterization of aerobic glucose metabolism in seven yeast species by 13C flux analysis and metabolomics. FEMS Yeast Res. 11, 263–72. https://doi.org/10.1111/j.1567-1364.2010.00713.x

Cordero Otero, R., Gaillardin, C., 1996. Efficient selection of hygromycin-B-resistant *Yarrowia lipolytica* transformants. Appl. Microbiol. Biotechnol. 46, 143–148. https://doi.org/10.1007/s002530050796

Cui, W., Wang, Q., Zhang, F., Zhang, S.C., Chi, Z.M., Madzak, C., 2011. Direct conversion of inulin into single cell protein by the engineered *Yarrowia lipolytica* carrying inulinase gene. Process Biochem. 46, 1442–1448. https://doi.org/10.1016/j.procbio.2011.03.017

Curran, K.A., Leavitt, J.M., Karim, A.S., Alper, H.S., 2013. Metabolic engineering of muconic acid production in *Saccharomyces cerevisiae*. Metab. Eng. 15, 55–66. https://doi.org/10.1016/j.ymben.2012.10.003

Curran, K.A., Morse, N.J., Markham, K.A., Wagman, A.M., Gupta, A., Alper, H.S., 2015. Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. ACS Synth. Biol. 4, 824–832. https://doi.org/10.1021/sb5003357

Czerwiec, Q., Idrissitaghki, A., Imatoukene, N., Nonus, M., Thomasset, B., Nicaud, J., Rossignol, T., 2019. Optimization of cyclopropane fatty acids production in *Yarrowia lipolytica*. Yeast 36, yea.3379. https://doi.org/10.1002/yea.3379

Darvishi, F., Ariana, M., Marella, E.R., Borodina, I., 2018. Advances in synthetic biology of oleaginous yeast *Yarrowia lipolytica* for producing non-native chemicals. Appl. Microbiol. Biotechnol. 102, 5925–5938. https://doi.org/10.1007/s00253-018-9099-x

Darvishi, F., Fathi, Z., Ariana, M., Moradi, H., 2017. *Yarrowia lipolytica* as a workhorse for biofuel production. Biochem. Eng. J. 127, 87–96. https://doi.org/10.1016/j.bej.2017.08.013

Dickinson, J.R., Salgado, L.E.J., Hewlins, M.J.E., 2003. The catabolism of amino acids to long chain and complex alcohols in *Saccharomyces cerevisiae*. J. Biol. Chem. 278, 8028–8034. https://doi.org/10.1074/jbc.M211914200

Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuvéglise, C., Talla, E., Goffard, N., Frangeul, L., Aigle, M., Anthouard, V., Babour, A., Barbe, V., Barnay, S., Blanchin, S., Beckerich, J.-M., Beyne, E., Bleykasten, C., Boisramé, A., Boyer, J., Cattolico, L., Confanioleri, F., De Daruvar, A., Despons, L., Fabre, E., Fairhead, C., Ferry-Dumazet, H., Groppi, A., Hantraye, F., Hennequin, C., Jauniaux, N., Joyet, P., Kachouri, R., Kerrest, A., Koszul, R., Lemaire, M., Lesur, I., Ma, L., Muller, H., Nicaud, J.-M., Nikolski, M., Oztas, S., Ozier-Kalogeropoulos, O., Pellenz, S., Potier, S., Richard,

G.-F., Straub, M.-L., Suleau, A., Swennen, D., Tekaia, F., Wésolowski-Louvel, M., Westhof, E., Wirth, B., Zeniou-Meyer, M., Zivanovic, I., Bolotin-Fukuhara, M., Thierry, A., Bouchier, C., Caudron, B., Scarpelli, C., Gaillardin, C., Weissenbach, J., Wincker, P., Souciet, J.-L., 2004. Genome evolution in yeasts. Nature 430, 35–44. https://doi.org/10.1038/nature02579

Durán, N., Justo, G.Z., Ferreira, C.V., Melo, P.S., Cordi, L., Martins, D., 2007. Violacein: properties and biological activities. Biotechnol. Appl. Biochem. 48, 127. https://doi.org/10.1042/ba20070115

Egermeier, M., Russmayer, H., Sauer, M., Marx, H., 2017. Metabolic Flexibility of *Yarrowia lipolytica* Growing on Glycerol. Front. Microbiol. 8. https://doi.org/10.3389/fmicb.2017.00049

Eichenberger, M., Lehka, B.J., Folly, C., Fischer, D., Martens, S., Simón, E., Naesby, M., 2017. Metabolic engineering of *Saccharomyces cerevisiae* for de novo production of dihydrochalcones with known antioxidant, antidiabetic, and sweet tasting properties. Metab. Eng. 39, 80–89. https://doi.org/10.1016/j.ymben.2016.10.019

ElObeid, A.S., Kamal-Eldin, A., Abdelhalim, M.A.K., Haseeb, A.M., 2017. Pharmacological Properties of Melanin and its Function in Health. Basic Clin. Pharmacol. Toxicol. 120, 515–522. https://doi.org/10.1111/bcpt.12748

Engler, C., Kandzia, R., Marillonnet, S., 2008. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. PLoS One 3, e3647. https://doi.org/10.1371/journal.pone.0003647

Etschmann, M., Bluemke, W., Sell, D., Schrader, J., 2002. Biotechnological production of 2-phenylethanol. Appl. Microbiol. Biotechnol. 59:1–8. https://doi.org/10.1007/s00253-002-0992-x

Etschmann, M.M.W., Schrader, J., 2006. An aqueous-organic two-phase bioprocess for efficient production of the natural aroma chemicals 2-phenylethanol and 2-phenylethylacetate with yeast. Appl. Microbiol. Biotechnol. 71, 440–443. https://doi.org/10.1007/s00253-005-0281-6

European parliament and Council 2008, n.d. REGULATION (EC) No 1334/2008 OF

THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 16 December 2008 on flavourings and certain food ingredients with flavouring properties for use in and on foods and amending Council Regulation (EEC) No 1601/91, Regulations.

Fickers, P., Benetti, P.-H., Waché, Y., Marty, A., Mauersberger, S., Smit, M.S., Nicaud, J.-M., 2005. Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. FEMS Yeast Res. 5, 527–543. https://doi.org/10.1016/j.femsyr.2004.09.004

Fickers, P., Le Dall, M.T., Gaillardin, C., Thonart, P., Nicaud, J.M., 2003. New disruption cassettes for rapid gene disruption and marker rescue in the yeast *Yarrowia lipolytica*. J. Microbiol. Methods 55, 727–737. doi: 10.1016/j.mimet.2003.07.003.

Finogenova, T. V., Morgunov, I.G., Kamzolova, S. V., Chernyavskaya, O.G., 2005. Organic acid production by the yeast *Yarrowia lipolytica*: A review of prospects. Appl. Biochem. Microbiol. 41, pages418–425. https://doi.org/10.1007/s10438-005-0076-7

Fournier, P., Abbas, A., Chasles, M., Kudla, B., Ogrydziak, D.M., Yaver, D., Xuan, J.W., Peito, A., Ribet, A.M., Feynerol, C., He, F., Gaillardin, C., 1993. Colocalization of centromeric and replicative functions on autonomously replicating sequences isolated from the yeast *Yarrowia lipolytica*. Proc. Natl. Acad. Sci. U. S. A. 90, 4912–4916. https://doi.org/10.1073/pnas.90.11.4912

Freije, C.A., Myhrvold, C., Boehm, C.K., Lin, A.E., Welch, N.L., Carter, A., Metsky, H.C., Luo, C.Y., Abudayyeh, O.O., Gootenberg, J.S., Yozwiak, N.L., Zhang, F., Sabeti, P.C., 2019. Programmable Inhibition and Detection of RNA Viruses Using Cas13. Mol. Cell 76, 826-837.e11. https://doi.org/10.1016/j.molcel.2019.09.013

Friedlander, J., Tsakraklides, V., Kamineni, A., Greenhagen, E.H., Consiglio, A.L., MacEwen, K., Crabtree, D. V, Afshar, J., Nugent, R.L., Hamilton, M.A., Joe Shaw, A., South, C.R., Stephanopoulos, G., Brevnova, E.E., 2016. Engineering of a high lipid producing *Yarrowia lipolytica* strain. Biotechnol. Biofuels 9, 77. https://doi.org/10.1186/s13068-016-0492-3

Future Market Insights, 2018. Products from Food Waste Market: Impact on Forecast and Analysis | Future Market Insights (FMI) [WWW Document]. Mark. Anal. Rep.

URL https://www.futuremarketinsights.com/reports/resveratrol-market.

Gamboa-Melendez, H., Larroude, M., Park, Y.K., Trebul, P., Nicaud, J.-M., Ledesma-Amaro, R., 2018. Synthetic Biology to Improve the Production of Lipases and Esterases (Review). Methods Mol. Biol. 1835, 229–242. https://doi.org/10.1007/978-1-4939-8672-9_13

Gao, C., Qi, Q., Madzak, C., Lin, C.S.K., 2015. Exploring medium-chain-length polyhydroxyalkanoates production in the engineered yeast *Yarrowia lipolytica*. J. Ind. Microbiol. Biotechnol. 42, 1255–1262. https://doi.org/10.1007/s10295-015-1649-y

Gao, D., Smith, S., Spagnuolo, M., Rodriguez, G., Blenner, M., 2018. Dual CRISPR-Cas9 Cleavage mediated Gene Excision and Targeted Integration in *Yarrowia lipolytica*. Biotechnol. J. e1700590. https://doi.org/10.1002/biot.201700590

Gao, M., Cao, M., Suástegui, M., Walker, J., Quiroz, N.R., Wu, Y., Tribby, D., Okerlund, A., Stanley, L., Shanks, J. V., Shao, Z., 2017. Innovating a nonconventional yeast platform for producing shikimate as the building block of high-value aromatics. ACS Synth. Biol. 6, 29–38. https://doi.org/10.1021/acssynbio.6b00132

Gao, S., Han, L., Zhu, L., Ge, M., Yang, S., Jiang, Y., Chen, D., 2014. One-step integration of multiple genes into the oleaginous yeast *Yarrowia lipolytica*. Biotechnol. Lett. 36, 2523–2528. https://doi.org/10.1007/s10529-014-1634-y

Gao, S., Tong, Y., Wen, Z., Zhu, L., Ge, M., Chen, D., Jiang, Y., Yang, S., 2016. Multiplex gene editing of the *Yarrowia lipolytica* genome using the CRISPR-Cas9 system. J. Ind. Microbiol. Biotechnol. 43, 1085–1093. https://doi.org/10.1007/s10295-016-1789-8

Gao, S., Tong, Y., Zhu, L., Ge, M., Zhang, Y., Chen, D., Jiang, Y., Yang, S., 2017. Iterative integration of multiple-copy pathway genes in *Yarrowia lipolytica* for heterologous β-carotene production. Metab. Eng. 41, 192–201. https://doi.org/10.1016/j.ymben.2017.04.004

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Iii, C.A.H., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat.

Methods 6, 343–345. https://doi.org/10.1038/nmeth.1318

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., Lim, W.A., Weissman, J.S., Qi, L.S., 2013. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell 154, 442–451. https://doi.org/10.1016/j.cell.2013.06.044

Gottardi, M., Reifenrath, M., Boles, E., Tripp, J., 2017. Pathway engineering for the production of heterologous aromatic chemicals and their derivatives in *Saccharomyces cerevisiae*: bioconversion from glucose. FEMS Yeast Res. https://doi.org/10.1093/femsyr/fox035

Grand View Research, 2016a. White Biotechnology Market Size | Industry Report, 2013-2024. Market Analysis Report. URL https://www.grandviewresearch.com/industry-analysis/white-biotechnology-market.

Grand View Research, 2016b. Flavonoids Market Size, Share | Global Industry Report, 2018-2025. Market Analysis Report URL https://www.grandviewresearch.com/industry-analysis/flavonoids-market.

Groenewald, M., Boekhout, T., Neuvéglise, C., Gaillardin, C., van Dijck, P.W.M., Wyss, M., 2014. *Yarrowia lipolytica*: safety assessment of an oleaginous yeast with a great industrial potential. Crit. Rev. Microbiol. 40, 187–206. https://doi.org/10.3109/1040841X.2013.770386

Guo, J., Rao, Z., Yang, T., Man, Z., Xu, M., Zhang, X., Yang, S.-T., 2015. Cloning and identification of a novel tyrosinase and its overexpression in *Streptomyces kathirae* SC-1 for enhancing melanin production. FEMS Microbiol. Lett. 362. https://doi.org/10.1093/femsle/fnv041

Haddouche, R., Delessert, S., Sabirova, J., Neuvéglise, C., Poirier, Y., Nicaud, J.-M., 2010. Roles of multiple acyl-CoA oxidases in the routing of carbon flow towards β-oxidation and polyhydroxyalkanoate biosynthesis in *Yarrowia lipolytica*. FEMS Yeast Res. 10, 917–927. https://doi.org/10.1111/j.1567-1364.2010.00670.x

Haddouche, R., Poirier, Y., Delessert, S., Sabirova, J., Pagot, Y., Neuvéglise, C.,

Nicaud, J.-M., 2011. Engineering polyhydroxyalkanoate content and monomer composition in the oleaginous yeast *Yarrowia lipolytica* by modifying the ß-oxidation multifunctional protein. Appl. Microbiol. Biotechnol. 91, 1327–1340. https://doi.org/10.1007/s00253-011-3331-2

Hamilton, M., Consiglio, A.L., MacEwen, K., Shaw, A.J., Tsakraklides, V., 2020. Identification of a *Yarrowia lipolytica* acetamidase and its use as a yeast genetic marker. Microb. Cell Fact. 19, 22. https://doi.org/10.1186/s12934-020-1292-9

Hanko, E.K.R., Denby, C.M., Sànchez I Nogué, V., Lin, W., Ramirez, K.J., Singer, C.A., Beckham, G.T., Keasling, J.D., 2018. Engineering β-oxidation in *Yarrowia lipolytica* for methyl ketone production. Metab. Eng. 48, 52–62. https://doi.org/10.1016/j.ymben.2018.05.018

Hansen, E.H., Møller, B.L., Kock, G.R., Bünner, C.M., Kristensen, C., Jensen, O.R., Okkels, F.T., Olsen, C.E., Motawia, M.S., Hansen, J., 2009. De novo biosynthesis of Vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). Appl. Environ. Microbiol. 75, 2765–2774. https://doi.org/10.1128/AEM.02681-08

Hartley, J.L., Temple, G.F., Brasch, M.A., 2000. DNA cloning using in vitro site-specific recombination. Genome Res. 10, 1788–1795. doi: 10.1101/gr.143000

Hartmann, M., Schneider, T.R., Pfeil, A., Heinrich, G., Lipscomb, W.N., Braus, G.H., 2003. Evolution of feedback-inhibited beta /alpha barrel isoenzymes by gene duplication and a single mutation. Proc. Natl. Acad. Sci. U. S. A. 100, 862–7. https://doi.org/10.1073/pnas.0337566100

Hassing, E.J., de Groot, P.A., Marquenie, V.R., Pronk, J.T., Daran, J.M.G., 2019. Connecting central carbon and aromatic amino acid metabolisms to improve de novo 2-phenylethanol production in *Saccharomyces cerevisiae*. Metab. Eng. 56, 165–180. https://doi.org/10.1016/j.ymben.2019.09.011

Hazelwood, L.A., Daran, J.M., Van Maris, A.J.A., Pronk, J.T., Dickinson, J.R., 2008. The Ehrlich pathway for fusel alcohol production: A century of research on *Saccharomyces cerevisiae* metabolism. Appl. Environ. Microbiol. 74(8), 2259–2266. https://doi.org/10.1128/AEM.02625-07

Heux, S., Meynial-Salles, I., O'Donohue, M.J., Dumon, C., 2015. White biotechnology: State of the art strategies for the development of biocatalysts for biorefining. Biotechnol. Adv. 33(8):1653-70 https://doi.org/10.1016/j.biotechadv.2015.08.004

Holkenbrink, C., Dam, M.I., Kildegaard, K.R., Beder, J., Dahlin, J., Doménech Belda, D., Borodina, I., 2018. EasyCloneYALI: CRISPR/Cas9-Based Synthetic Toolbox for Engineering of the Yeast *Yarrowia lipolytica*. Biotechnol. J. 13, e1700543. https://doi.org/10.1002/biot.201700543

Holt, S., Miks, M.H., De Carvalho, B.T., Foulquié-Moreno, M.R., Thevelein, J.M., 2019. The molecular biology of fruity and floral aromas in beer and other alcoholic beverages. FEMS Microbiol. Rev. https://doi.org/10.1093/femsre/fuy041

Hua, D., Xu, P., 2011. Recent advances in biotechnological production of 2-phenylethanol. Biotechnol. Adv. 29(6):654-60 https://doi.org/10.1016/j.biotechadv.2011.05.001

Huang, L.L., Xue, Z., Zhu, Q.Q., 2006. Method for the production of resveratrol in a recombinant oleaginous microorganism. United States Patent Application Publication. Pub. N°: US 2009/0082286.

Huccetogullari, D., Luo, Z.W., Lee, S.Y., 2019. Metabolic engineering of microorganisms for production of aromatic compounds. Microb. Cell Fact. 18, 41. https://doi.org/10.1186/s12934-019-1090-4

Imatoukene, N., Verbeke, J., Beopoulos, A., Idrissi Taghki, A., Thomasset, B., Sarde, C.O., Nonus, M., Nicaud, J.M., 2017. A metabolic engineering strategy for producing conjugated linoleic acids using the oleaginous yeast *Yarrowia lipolytica*. Appl. Microbiol. Biotechnol. 101, 4605–4616. https://doi.org/10.1007/s00253-017-8240-6

Iraqui, I., Vissers, S., André, B., Urrestarazu, A., 1999. Transcriptional Induction by Aromatic Amino Acids in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 19, 3360–3371. doi: 10.1128/mcb.19.5.3360

Iraqui, I., Vissers, S., Cartiaux, M., Urrestarazu, A., 1998. Characterisation of *Saccharomyces cerevisiae* ARO8 and ARO9 genes encoding aromatic aminotransferases I and II reveals a new aminotransferase subfamily. Mol. Gen.

Genet. 257, 238–248. https://doi.org/10.1007/s004380050644

Ito, H., Inouhe, M., Tohoyama, H., Joho, M., 2007. Characteristics of copper tolerance in *Yarrowia lipolytica*. BioMetals 20, 773–780. https://doi.org/10.1007/s10534-006-9040-0

Jiang, H., Morgan, J.A., 2004. Optimization of an in Vivo Plant P450 Monooxygenase System in *Saccharomyces cerevisiae*. Biotechnol. Bioeng. 85, 130–137. https://doi.org/10.1002/bit.10867

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E., 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science (80-.). 337, 816–821. https://doi.org/10.1126/science.1225829

Johnravindar, D., Karthikeyan, O.P., Selvam, A., Murugesan, K., Wong, J.W.C., 2018. Lipid accumulation potential of oleaginous yeasts: A comparative evaluation using food waste leachate as a substrate. Bioresour. Technol. 248, 221–228. https://doi.org/10.1016/j.biortech.2017.06.151

Juretzek, T., Dall, M.-T. Le, Mauersberger, S., Gaillardin, C., Barth, G., Nicaud, J.-M., 2001. Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. Yeast 18, 97–113. https://doi.org/10.1002/1097-0061(20010130)18:2<97::AID-YEA652>3.0.CO;2-U

Kerkhoven, E.J., Pomraning, K.R., Baker, S.E., Nielsen, J., 2016. Regulation of amino-acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*. npj Syst. Biol. Appl. 2, npjsba20165. https://doi.org/10.1038/npjsba.2016.5

Kerscher, S., Durstewitz, G., Casaregola, S., Gaillardin, C., Brandt, U., 2001. The complete mitochondrial genome of *Yarrowia lipolytica*. Comp. Funct. Genomics 2, 80–90. https://doi.org/10.1002/cfg.72

Khalil, A.S., Collins, J.J., 2010. Synthetic biology: Applications come of age. Nat. Rev. Genet. 11, 367–379. https://doi.org/10.1038/nrg2775

Kholany, M., Trébulle, P., Martins, M., Ventura, S.P., Nicaud, J., Coutinho, J.A., 2019. Extraction and purification of violacein from *Yarrowia lipolytica* cells using aqueous solutions of surfactants. J. Chem. Technol. Biotechnol. 95: 1126–1134. https://doi.org/10.1002/jctb.6297

Kim, Y.J., Wu, W., Chun, S.E., Whitacre, J.F., Bettinger, C.J., 2013. Biologically derived melanin electrodes in aqueous sodium-ion energy storage devices. Proc. Natl. Acad. Sci. U. S. A. 110, 20912–20917. https://doi.org/10.1073/pnas.1314345110

Kiran, G.S., Dhasayan, A., Lipton, A.N., Selvin, J., Arasu, M. V., Al-Dhabi, N.A., 2014. Melanin-templated rapid synthesis of silver nanostructures. J. Nanobiotechnology 12, 18. https://doi.org/10.1186/1477-3155-12-18

Knight, T., 2003. Idempotent Vector Design for Standard Assembly of Biobricks. Tech. rep., MIT Synthetic Biology Working Group Technical Reports 2003. [http://hdl.handle.net/1721.1/21168]

Koonin, E. V, Makarova, K.S., Zhang, F., 2017. Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. 37, 67–78. https://doi.org/10.1016/j.mib.2017.05.008

Koopman, F., Beekwilder, J., Crimi, B., van Houwelingen, A., Hall, R.D., Bosch, D., van Maris, A.J., Pronk, J.T., Daran, J.-M., 2012. De novo production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*. Microb. Cell Fact. 11, 155. https://doi.org/10.1186/1475-2859-11-155

Krappmann, S., Lipscomb, W.N., Braus, G.H., 2000. Coevolution of transcriptional and allosteric regulation at the chorismate metabolic branch point of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. U. S. A. 97, 13585–13590. https://doi.org/10.1073/pnas.240469697

Kretzschmar, A., Otto, C., Holz, M., Werner, S., Hübner, L., Barth, G., 2013. Increased homologous integration frequency in *Yarrowia lipolytica* strains defective in non-homologous end-joining. Curr. Genet. 59, 63–72. https://doi.org/10.1007/s00294-013-0389-7

Kuepper, J., Dickler, J., Biggel, M., Behnken, S., Jäger, G., Wierckx, N., Blank, L.M., 2015. Metabolic engineering of Pseudomonas putida KT2440 to produce anthranilate from glucose. Front. Microbiol. 6, 1310. https://doi.org/10.3389/fmicb.2015.01310

Larroude, M., Celinska, E., Back, A., Thomas, S., Nicaud, J.-M., Ledesma-Amaro, R.,

2017. A synthetic biology approach to transform *Yarrowia lipolytica* into a competitive biotechnological producer of β-carotene. Biotechnol. Bioeng. 115:464–472. https://doi.org/10.1002/bit.26473

Larroude, M., Park, Y.K., Soudier, P., Kubiak, M., Nicaud, J.M., Rossignol, T., 2019. A modular Golden Gate toolkit for *Yarrowia lipolytica* synthetic biology. Microb. Biotechnol. 12(6), 1249–1259 https://doi.org/10.1111/1751-7915.13427

Larroude, M., Rossignol, T., Nicaud, J.-M., Ledesma-Amaro, R., 2018. Synthetic biology tools for engineering *Yarrowia lipolytica*. Biotechnol. Adv. 36, 2150–2164. https://doi.org/10.1016/J.BIOTECHADV.2018.10.004

Larroude, M., Trabelsi, H., Nicaud, J.-M., Rossignol, T., 2020. A set of *Yarrowia lipolytica* CRISPR/Cas9 vectors for exploiting wild-type strain diversity. Biotechnol. Lett. 42, 773–785 https://doi.org/10.1007/s10529-020-02805-4

Lazar, Z., Rossignol, T., Verbeke, J., Coq, A.-M.C.-L., Nicaud, J.-M., Robak, M., 2013. Optimized invertase expression and secretion cassette for improving *Yarrowia lipolytica* growth on sucrose for industrial applications. J. Ind. Microbiol. Biotechnol. 40, 1273–1283. https://doi.org/10.1007/s10295-013-1323-1

Ledesma-Amaro, R., Nicaud, J.-M., 2016a. Metabolic Engineering for Expanding the Substrate Range of *Yarrowia lipolytica*. Trends Biotechnol. 34, 798–809. https://doi.org/10.1016/j.tibtech.2016.04.010

Ledesma-Amaro, R., Nicaud, J.-M., 2016b. *Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids. Prog. Lipid Res. 61, 40–50. https://doi.org/10.1016/j.plipres.2015.12.001

Lehninger, A.L., Nelson, D.L.., Cox, M.M., 2000. Lehninger Principles of Biochemistry, 7th ed. New York: Worth Publishers.

Leonard, E., Yan, Y., Fowler, Z.L., Li, Z., Lim, C.G., Lim, K.H., Koffas, M.A.G., 2008. Strain improvement of recombinant *Escherichia coli* for efficient production of plant flavonoids. Mol. Pharm. 5, 257–265. https://doi.org/10.1021/mp7001472

Leplat, C., Nicaud, J.-M., Rossignol, T., 2018. Overexpression screen reveals transcription factors involved in lipid accumulation in *Yarrowia lipolytica*. FEMS Yeast Res. 18. https://doi.org/10.1093/femsyr/foy037

Leplat, C., Nicaud, J.-M., Rossignol, T., 2015. High-throughput transformation method for *Yarrowia lipolytica* mutant library screening. FEMS Yeast Res. 15. https://doi.org/10.1093/femsyr/fov052

Li, Z.-J., Qiao, K., Liu, N., Stephanopoulos, G., 2017. Engineering *Yarrowia lipolytica* for poly-3-hydroxybutyrate production. J. Ind. Microbiol. Biotechnol. 44, 605–612. https://doi.org/10.1007/s10295-016-1864-1

Lim, C.G., Fowler, Z.L., Hueller, T., Schaffer, S., Koffas, M.A.G., 2011. High-yield resveratrol production in engineered *Escherichia coli*. Appl. Environ. Microbiol. 77, 3451–3460. https://doi.org/10.1128/AEM.02186-10

Liu, H.-H., Ji, X.-J., Huang, H., 2015. Biotechnological applications of *Yarrowia lipolytica*: Past, present and future. Biotechnol. Adv. 33, 1522–1546. https://doi.org/10.1016/j.biotechadv.2015.07.010

Liu, N., Qiao, K., Stephanopoulos, G., 2016. 13C Metabolic Flux Analysis of acetate conversion to lipids by *Yarrowia lipolytica*. Metab. Eng. 38, 86–97. https://doi.org/10.1016/j.ymben.2016.06.006

Liu, Q., Yu, T., Li, X., Chen, Yu, Campbell, K., Nielsen, J., Chen, Yun, 2019. Rewiring carbon metabolism in yeast for high level production of aromatic chemicals. Nat. Commun. 10, 1–13. https://doi.org/10.1038/s41467-019-12961-5

Liu, Y.C., Chen, S.M., Liu, J.H., Hsu, H.W., Lin, H.Y., Chen, S.Y., 2015. Mechanical and photo-fragmentation processes for nanonization of melanin to improve its efficacy in protecting cells from reactive oxygen species stress. J. Appl. Phys. 117, 064701. https://doi.org/10.1063/1.4907997

Ljungdahl, P.O., Daignan-Fornier, B., 2012. Regulation of amino acid, nucleotide, and phosphate metabolism in *Saccharomyces cerevisiae*. Genetics. 190(3), 885-929. https://doi.org/10.1534/genetics.111.133306

Lustig, A.J., 1999. The kudos of non-homologous end-joining. Nat. Genet. 23, pages130–131. https://doi.org/10.1038/13755

Luttik, M.A.H., Vuralhan, Z., Suir, E., Braus, G.H., Pronk, J.T., Daran, J.M., 2008. Alleviation of feedback inhibition in *Saccharomyces cerevisiae* aromatic amino acid biosynthesis: Quantification of metabolic impact. Metab. Eng. 10, 141–153.

https://doi.org/10.1016/j.ymben.2008.02.002

Lv, Y., Edwards, H., Zhou, J., Xu, P., 2019a. Combining 26s rDNA and the Cre-loxP System for Iterative Gene Integration and Efficient Marker Curation in *Yarrowia lipolytica*. ACS Synth. Biol. 8, 568–576. https://doi.org/10.1021/acssynbio.8b00535

Lv, Y., Marsafari, M., Koffas, M.A.G., Zhou, J., Xu, P., 2019b. Optimizing oleaginous yeast cell factories for flavonoids and hydroxylated flavonoids biosynthesis. ACS Synth. Biol. 8, 11, 2514–2523. https://doi.org/10.1021/acssynbio.9b00193

Lyu, X., Ng, K.R., Mark, R., Lee, J.L., Chen, W.N., 2018. Comparative metabolic profiling of engineered *Saccharomyces cerevisiae* with enhanced flavonoids production. J. Funct. Foods 44, 274–282. https://doi.org/10.1016/j.jff.2018.03.012

Madzak, C., 2015. *Yarrowia lipolytica*: recent achievements in heterologous protein expression and pathway engineering. Appl. Microbiol. Biotechnol. 99, 4559–4577. https://doi.org/10.1007/s00253-015-6624-z

Madzak, C., Blanchin-Roland, S., Cordero Otero, R.R., Gaillardin, C., 1999. Functional analysis of upstream regulating regions from the *Yarrowia lipolytica* XPR2 promoter. Microbiology 145 (1), 75–87. https://doi.org/10.1099/13500872-145-1-75

Madzak, C., Tréton, B., Blanchin-Roland, S., 2000. Strong hybrid promoters and integrative expression/secretion vectors for quasi-constitutive expression of heterologous proteins in the yeast *Yarrowia lipolytica*. J. Mol. Microbiol. Biotechnol. 2, 207–216.

Magdouli, S., Guedri, T., Tarek, R., Brar, S.K., Blais, J.F., 2017. Valorization of raw glycerol and crustacean waste into value added products by *Yarrowia lipolytica*. Bioresour. Technol. 243, 57–68. https://doi.org/10.1016/j.biortech.2017.06.074

Magnan, C., Yu, J., Chang, I., Jahn, E., Kanomata, Y., Wu, J., Zeller, M., Oakes, M., Baldi, P., Sandmeyer, S., 2016. Sequence Assembly of *Yarrowia lipolytica* Strain W29/CLIB89 Shows Transposable Element Diversity. PLoS One 11, e0162363. https://doi.org/10.1371/journal.pone.0162363

Mahmoud, Y.A.G., 2004. Uptake of Radionuclides by Some Fungi. Mycobiology 32,

110. https://doi.org/10.4489/myco.2004.32.3.110

Mahr, R., Frunzke, J., 2016. Transcription factor-based biosensors in biotechnology: current state and future prospects. Appl. Microbiol. Biotechnol. 100(1):79-90 https://doi.org/10.1007/s00253-015-7090-3

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.J., Charpentier, E., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Terns, R.M., Terns, M.P., White, M.F., Yakunin, A.F., Garrett, R.A., Van Der Oost, J., Backofen, R., Koonin, E. V., 2015. An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. 13, 722–736. https://doi.org/10.1038/nrmicro3569

Mali, P., Esvelt, K.M., Church, G.M., 2013. Cas9 as a versatile tool for engineering biology. Nat. Methods 10, 957–963. https://doi.org/10.1038/nmeth.2649

Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G.A., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J.A., Daran, J.-M.G., 2015. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. FEMS Yeast Res. 15. https://doi.org/10.1093/femsyr/fov004

Mao, J., Liu, Q., Song, X., Wang, H., Feng, H., Xu, H., Qiao, M., 2017. Combinatorial analysis of enzymatic bottlenecks of l-tyrosine pathway by p-coumaric acid production in *Saccharomyces cerevisiae*. Biotechnol. Lett. 39, 977–982. https://doi.org/10.1007/s10529-017-2322-5

Markham, K.A., Alper, H.S., 2018. Synthetic Biology Expands the Industrial Potential of *Yarrowia lipolytica*. Trends Biotechnol. 36(10), 1085-1095 https://doi.org/10.1016/j.tibtech.2018.05.004

Martínez, L.M., Martinez, A., Gosset, G., 2019. Production of Melanins With Recombinant Microorganisms. Front. Bioeng. Biotechnol. 7. https://doi.org/10.3389/fbioe.2019.00285

Matthäus, F., Ketelhot, M., Gatter, M., Barth, G., 2014. Production of Lycopene in the Non-Carotenoid-Producing Yeast *Yarrowia lipolytica*. Appl. Environ. Microbiol. 80, 1660–1669. https://doi.org/10.1128/AEM.03167-13

Matthews, D.E., 2007. An Overview of Phenylalanine and Tyrosine Kinetics in Humans. J Nutr. 137(6 Suppl 1): 1549S–1575S. doi: 10.1093/jn/137.6.1549S

Mauersberger, S., Wang, H.J., Gaillardin, C., Barth, G., Nicaud, J.M., 2001. Insertional mutagenesis in the n-alkane-assimilating yeast *Yarrowia lipolytica*: generation of tagged mutations in genes involved in hydrophobic substrate utilization. J. Bacteriol. 183, 5102–5109.

Mei, J., Min, H., Lü, Z., 2009. Enhanced biotransformation of l-phenylalanine to 2-phenylethanol using an in situ product adsorption technique. Process Biochem. 44, 886–890. https://doi.org/10.1016/j.procbio.2009.04.012

Mei, Y.Z., Liu, R.X., Wang, D.P., Wang, X., Dai, C.C., 2015. Biocatalysis and biotransformation of resveratrol in microorganisms. Biotechnol. Lett. 37, pages9–18. https://doi.org/10.1007/s10529-014-1651-x

Miller, K.K., Alper, H.S., 2019. *Yarrowia lipolytica*: more than an oleaginous workhorse. Appl. Microbiol. Biotechnol. 1–12. https://doi.org/10.1007/s00253-019-10200-x

Mirończuk, A.M., Rzechonek, D.A., Biegalska, A., Rakicka, M., Dobrowolski, A., 2016. A novel strain of *Yarrowia lipolytica* as a platform for value-added product synthesis from glycerol. Biotechnol. Biofuels 9, 180. https://doi.org/10.1186/s13068-016-0593-z

Morin, N., Cescut, J., Beopoulos, A., Lelandais, G., Le Berre, V., Uribelarrea, J.L., Molina-Jouve, C., Nicaud, J.M., 2011. Transcriptomic analyses during the transition from biomass production to lipid accumulation in the oleaginous yeast *Yarrowia lipolytica*. PLoS One 6. https://doi.org/10.1371/journal.pone.0027966

Morse, N.J., Wagner, J.M., Reed, K.B., Gopal, M.R., Lauffer, L.H., Alper, H.S., 2018. T7 Polymerase Expression of Guide RNAs in vivo Allows Exportable CRISPR-Cas9 Editing in Multiple Yeast Hosts. ACS Synth. Biol. 7, 1075–1084. https://doi.org/10.1021/acssynbio.7b00461

Mosst, A.R., Schoenheimer, R., 1940. The conversion of phenylalanine to tyrosine in normal rats. J. Biol. Chem. 135, 415-429

Müller, S., Sandal, T., Kamp-Hansen, P., Dalbøge, H., 1998. Comparison of expression

systems in the yeasts *Saccharomyces cerevisiae*, *Hansenula polymorpha*, *Klyveromyces lactis*, *Schizosaccharomyces pombe* and *Yarrowia lipolytica*. Cloning of two novel promoters from *Yarrowia lipolytica*. Yeast 14, 1267–1283. https://doi.org/10.1002/(SICI)1097-0061(1998100)14:14<1267::AID-YEA327>3.0.CO;2-2

Naumova, E., Naumov, G., Fournier, P., Nguyen, H. V, Gaillardin, C., 1993. Chromosomal polymorphism of the yeast *Yarrowia lipolytica* and related species: electrophoretic karyotyping and hybridization with cloned genes. Curr. Genet. 23, 450–454. doi: 10.1007/BF00312633

Nicaud, J.-M., 2012. *Yarrowia lipolytica*. Yeast 29, 409–418. https://doi.org/10.1002/yea.2921

Nicaud, J.-M., Fabre, E., Gaillardin, C., 1989. Expression of invertase activity in *Yarrowia lipolytica* and its use as a selective marker. Curr. Genet. 16, 253–260. https://doi.org/10.1007/BF00422111

Nicaud, J.-M., Madzak, C., van den Broek, P., Gysler, C., Duboc, P., Niederberger, P., Gaillardin, C., 2002. Protein expression and secretion in the yeast *Yarrowia lipolytica*. FEMS Yeast Res. 2, 371–379. doi: 10.1016/S1567-1356(02)00082-X

Nicaud, J.M., Fournier, P., La Bonnardière, C., Chasles, M., Gaillardin, C., 1991. Use of ars18 based vectors to increase protein production in *Yarrowia lipolytica*. J. Biotechnol. 19, 259–270. https://doi.org/10.1016/0168-1656(91)9006 https://doi.org/10.1016/0168-1656(91)90063-2 3-2

Niehus, X., Crutz-Le Coq, A.M., Sandoval, G., Nicaud, J.M., Ledesma-Amaro, R., 2018. Engineering *Yarrowia lipolytica* to enhance lipid production from lignocellulosic materials. Biotechnol. Biofuels 11, 11. https://doi.org/10.1186/s13068-018-1010-6

Nielsen, J., Fussenegger, M., Keasling, J., Lee, S.Y., Liao, J.C., Prather, K., Palsson, B., 2014. Engineering synergy in biotechnology. Nat. Chem. Biol. 10(5):319-22 https://doi.org/10.1038/nchembio.1519

Nielsen, J., Keasling, J.D., 2016. Engineering Cellular Metabolism. Cell 164, 1185–1197. https://doi.org/10.1016/j.cell.2016.02.004

Nijveldt, R.J., Van Nood, E., Van Hoorn, D.E.C., Boelens, P.G., Van Norren, K., Van Leeuwen, P.A.M., 2001. Flavonoids: A review of probable mechanisms of action and potential applications. Am. J. Clin. Nutr. 74(4): 418–425 https://doi.org/10.1093/ajcn/74.4.418

Niu, W., Draths, K.M., Frost, J.W., 2002. Benzene-free synthesis of adipic acid. Biotechnol. Prog. 18, 201–211. https://doi.org/10.1021/bp010179x

Noda, S., Kondo, A., 2017. Recent Advances in Microbial Production of Aromatic Chemicals and Derivatives. Trends Biotechnol. 35(8):785-796 https://doi.org/10.1016/j.tibtech.2017.05.006

Nthangeni, M.B., Urban, P., Pompon, D., Smit, M.S., Nicaud, J.M., 2004. The use of *Yarrowia lipolytica* for the expression of human cytochrome P450 CYP1A1. Yeast 21, 583–592. https://doi.org/10.1002/yea.1127

Otto, C., Holz, M., Barth, G., 2013. Production of Organic Acids by *Yarrowia lipolytica*. Springer, Berlin, Heidelberg, pp. 137–149. https://doi.org/10.1007/978-3-642-38583-4_5

Overhage, J., Steinbüchel, A., Priefert, H., 2003. Highly Efficient Biotransformation of Eugenol to Ferulic Acid and Further Conversion to Vanillin in Recombinant Strains of *Escherichia coli*. Appl. Environ. Microbiol. 69, 6569–6576. https://doi.org/10.1128/AEM.69.11.6569-6576.2003

Palmer, C.M., Miller, K.K., Nguyen, A., Alper, H.S., 2020. Engineering 4-coumaroyl-CoA derived polyketide production in *Yarrowia lipolytica* through a β-oxidation mediated strategy. Metab. Eng. 57, 174–181. https://doi.org/10.1016/J.YMBEN.2019.11.006

Pandal, N., 2014. Global markets for flavors and fragrances. CHM034D. Market Analysis Report. BBC Research. https://www.bccresearch.com/market-research/chemicals/flavors-fragrances-markets-report.html

Pandey, R.P., Parajuli, P., Koffas, M.A.G., Sohng, J.K., 2016. Microbial production of natural and non-natural flavonoids: Pathway engineering, directed evolution and systems/synthetic biology. Biotechnol. Adv. https://doi.org/10.1016/j.biotechadv.2016.02.012

Pangeni, R., Sahni, J.K., Ali, J., Sharma, S., Baboota, S., 2014. Resveratrol: Review on therapeutic potential and recent advances in drug delivery. Expert Opin. Drug Deliv. https://doi.org/10.1517/17425247.2014.919253

Park, Y.-K., Dulermo, T., Ledesma-Amaro, R., Nicaud, J.-M., 2018. Optimization of odd chain fatty acid production by *Yarrowia lipolytica*. Biotechnol. Biofuels 11, 158. https://doi.org/10.1186/s13068-018-1154-4

Park, Y.K., Korpys, P., Kubiak, M., Celinska, E., Soudier, P., Trébulle, P., Larroude, M., Rossignol, T., Nicaud, J.M., 2019. Engineering the architecture of erythritol-inducible promoters for regulated and enhanced gene expression in *Yarrowia lipolytica*. FEMS Yeast Res. 19. https://doi.org/10.1093/femsyr/foy105

Pignède, G., Wang, H.J., Fudalej, F., Seman, M., Gaillardin, C., Nicaud, J.M., 2000. Autocloning and amplification of LIP2 in *Yarrowia lipolytica*. Appl. Environ. Microbiol. 66, 3283–3289. doi: 10.1128/aem.66.8.3283-3289.2000

Plonka, P.M., Grabacka, M., 2006. Melanin synthesis in microorganisms - Biotechnological and medical aspects. Acta Biochim. Pol. 53(3):429-443

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., Lim, W.A., 2013. Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. Cell 152, 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022

Rakicka, M., Rywińska, A., Cybulski, K., Rymowicz, W., 2016. Enhanced production of erythritol and mannitol by *Yarrowia lipolytica* in media containing surfactants. Brazilian J. Microbiol. 47, 417–423. https://doi.org/10.1016/j.bjm.2016.01.011

Reifenrath, M., Bauer, M., Oreb, M., Boles, E., 2018. Bacterial bifunctional chorismate mutase-prephenate dehydratase PheA increases flux into the yeast phenylalanine pathway and improves mandelic acid production. Metab. Eng. Commun. 7. https://doi.org/10.1016/j.mec.2018.e00079

Richard, G.-F., Kerrest, A., Lafontaine, I., Dujon, B., 2005. Comparative genomics of hemiascomycete yeasts: genes involved in DNA replication, repair, and recombination. Mol. Biol. Evol. 22, 1011–1023. https://doi.org/10.1093/molbev/msi083

Rigouin, C., Gueroult, M., Croux, C., Dubois, G., Borsenberger, V., Barbe, S., Marty, A., Daboussi, F., André, I., Bordes, F., 2017. Production of Medium Chain Fatty Acids by *Yarrowia lipolytica*: Combining Molecular Design and TALEN to Engineer the Fatty Acid Synthase. ACS Synth. Biol. 6, 1870–1879. https://doi.org/10.1021/acssynbio.7b00034

Rigouin, C., Lajus, S., Ocando, C., Borsenberger, V., Nicaud, J.M., Marty, A., Avérous, L., Bordes, F., 2019. Production and characterization of two medium-chain-length polydroxyalkanoates by engineered strains of *Yarrowia lipolytica*. Microb. Cell Fact. 18, 99. https://doi.org/10.1186/s12934-019-1140-y

Rodrigues, A.L., Trachtmann, N., Becker, J., Lohanatha, A.F., Blotenberg, J., Bolten, C.J., Korneli, C., de Souza Lima, A.O., Porto, L.M., Sprenger, G.A., Wittmann, C., 2013. Systems metabolic engineering of *Escherichia coli* for production of the antitumor drugs violacein and deoxyviolacein. Metab. Eng. 20, 29–41. https://doi.org/10.1016/j.ymben.2013.08.004

Rodriguez, A., Kildegaard, K.R., Li, M., Borodina, I., Nielsen, J., 2015. Establishment of a yeast platform strain for production of p-coumaric acid through metabolic engineering of aromatic amino acid biosynthesis. Metab. Eng. 31, 181–188. https://doi.org/10.1016/j.ymben.2015.08.003

Rodriguez, G.M., Hussain, M.S., Gambill, L., Gao, D., Yaguchi, A., Blenner, M., 2016. Engineering xylose utilization in *Yarrowia lipolytica* by understanding its cryptic xylose pathway. Biotechnol. Biofuels 9, 1–15. https://doi.org/10.1186/s13068-016-0562-6

Sánchez, B.J., Zhang, C., Nilsson, A., Lahtvee, P.-J., Kerkhoven, E.J., Nielsen, J., 2017. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol. Syst. Biol. 13, 935. https://doi.org/10.15252/msb.20167411

Sánchez, C., Braña, A.F., Méndez, C., Salas, J.A., 2006. Reevaluation of the violacein biosynthetic pathway and its relationship to indolocarbazole biosynthesis. ChemBioChem 7, 1231–1240. https://doi.org/10.1002/cbic.200600029

Schmaler-Ripcke, J., Sugareva, V., Gebhardt, P., Winkler, R., Kniemeyer, O.,

Heinekamp, T., Brakhage, A.A., 2009. Production of pyomelanin, a second type of melanin, via the tyrosine degradation pathway in *Aspergillus fumigatus*. Appl. Environ. Microbiol. 75, 493–503. https://doi.org/10.1128/AEM.02077-08

Schmid-Berger, N., Schmid, B., Barth, G., 1994. Ylt1, a highly repetitive retrotransposon in the genome of the dimorphic fungus *Yarrowia lipolytica*. J. Bacteriol. 176, 2477–2482.

Schnappauf, G., Lipscomb, W.N., Braus, G.H., 1998. Separation of inhibition and activation of the allosteric yeast chorismate mutase. Proc. Natl. Acad. Sci. U. S. A. 95, 2868–2873. https://doi.org/10.1073/pnas.95.6.2868

Schwartz, C., Cheng, J.-F., Evans, R., Schwartz, C.A., Wagner, J.M., Anglin, S., Beitz, A., Pan, W., Lonardi, S., Blenner, M., Alper, H.S., Yoshikuni, Y., Wheeldon, I., 2019. Validating genome-wide CRISPR-Cas9 function improves screening in the oleaginous yeast *Yarrowia lipolytica*. Metab. Eng. 55, 102–110. https://doi.org/10.1016/j.ymben.2019.06.007

Schwartz, C., Curtis, N., Löbs, A.K., Wheeldon, I., 2018. Multiplexed CRISPR activation of cryptic sugar metabolism enables *Yarrowia lipolytica* growth on cellobiose. Biotechnol. J. 13, e1700584. https://doi.org/10.1002/biot.201700584

Schwartz, C., Frogue, K., Ramesh, A., Misa, J., Wheeldon, I., 2017a. CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in *Yarrowia lipolytica*. Biotechnol. Bioeng. 114, 2896–2906. https://doi.org/10.1002/bit.26404

Schwartz, C., Shabbir-Hussain, M., Frogue, K., Blenner, M., Wheeldon, I., 2017b. Standardized Markerless Gene Integration for Pathway Engineering in *Yarrowia lipolytica*. ACS Synth. Biol. 6, 402–409. https://doi.org/10.1021/acssynbio.6b00285

Schwartz, C.M., Hussain, M.S., Blenner, M., Wheeldon, I., 2016. Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR–Cas9-Mediated Genome Editing in *Yarrowia lipolytica*. ACS Synth. Biol. 5, 356–359. https://doi.org/10.1021/acssynbio.5b00162

Shang, Y., Wei, W., Zhang, P., Ye, B.-C., 2020. Engineering *Yarrowia lipolytica* for

Enhanced Production of Arbutin. J. Agric. Food Chem. 68, 1364–1372. https://doi.org/10.1021/acs.jafc.9b07151

Solano, F., 2017. Melanin and melanin-related polymers as materials with biomedical and biotechnological applications— Cuttlefish ink and mussel foot proteins as inspired biomolecules. Int. J. Mol. Sci. 18(7):1561 https://doi.org/10.3390/ijms18071561

Soudier, P., Larroude, M., Celińska, E., Rossignol, T., Nicaud, J.-M., 2019. Selection of Heterologous Protein-Producing Strains in *Yarrowia lipolytica*, in: Methods in Molecular Biology (Clifton, N.J.). pp. 153–168. https://doi.org/10.1007/978-1-4939-9024-5_6

Spagnuolo, M., Shabbir Hussain, M., Gambill, L., Blenner, M., 2018. Alternative Substrate Metabolism in *Yarrowia lipolytica*. Front. Microbiol. 9. https://doi.org/10.3389/fmicb.2018.01077

Stephanopoulos, G., 2012. Synthetic biology and metabolic engineering. ACS Synth. Biol. 1, 514–525. https://doi.org/10.1021/sb300094q

Suástegui, M., Guo, W., Feng, X., Shao, Z., 2016. Investigating strain dependency in the production of aromatic compounds in *Saccharomyces cerevisiae*. Biotechnol. Bioeng. 113, 2676–2685. https://doi.org/10.1002/bit.26037

Suástegui, M., Shao, Z., 2016. Yeast factories for the production of aromatic compounds: from building blocks to plant secondary metabolites. J. Ind. Microbiol. Biotechnol. 43, 1611–1624. https://doi.org/10.1007/s10295-016-1824-9

Suástegui, M., Yu Ng, C., Chowdhury, A., Sun, W., Cao, M., House, E., Maranas, C.D., Shao, Z., 2017. Multilevel engineering of the upstream module of aromatic amino acid biosynthesis in *Saccharomyces cerevisiae* for high production of polymer and drug precursors. Metab. Eng. 42, 134–144. https://doi.org/10.1016/j.ymben.2017.06.008

Sun, H., Zhao, D., Xiong, B., Zhang, C., Bi, C., 2016. Engineering *Corynebacterium glutamicum* for violacein hyper production. Microb. Cell Fact. 15, 148. https://doi.org/10.1186/s12934-016-0545-0

Swiat, M.A., Dashko, S., den Ridder, M., Wijsman, M., van der Oost, J., Daran, J.-M.,

Daran-Lapujade, P., 2017. FnCpf1: a novel and efficient genome editing tool for *Saccharomyces cerevisiae*. Nucleic Acids Res. 45, 12585–12598. https://doi.org/10.1093/nar/gkx1007

Thapa, S.B., Pandey, R.P., Park, Y. Il, Kyung Sohng, J.K., 2019. Biotechnological Advances in Resveratrol Production and its Chemical Diversity. Molecules 24, 2571. https://doi.org/10.3390/molecules24142571

Thevenieau, F., Nicaud, J.-M., Gaillardin, C., 2009. Applications of the Non-Conventional Yeast *Yarrowia lipolytica*, in: Satyanarayana, T., Kunze, G. (Eds.), Yeast Biotechnology: Diversity and Applications. Springer Netherlands, pp. 589–613.

Thorne, R.S.W., 1949. Nitrogen metabolism of yeast. a consideration of the mode of assimilation of amino acids. J. Inst. Brew. 55, 201–222. https://doi.org/10.1002/j.2050-0416.1949.tb01492.x

Trabelsi, H., Koch, M., Faulon, J.L., 2018. Building a minimal and generalizable model of transcription factor–based biosensors: Showcasing flavonoids. Biotechnol. Bioeng. 115, 2292–2304. https://doi.org/10.1002/bit.26726

Trassaert, M., Vandermies, M., Carly, F., Denies, O., Thomas, S., Fickers, P., Nicaud, J.-M., 2017. New inducible promoter for gene expression and synthetic biology in *Yarrowia lipolytica*. Microb. Cell Fact. 16, 141. https://doi.org/10.1186/s12934-017-0755-0

Trebulle, P., 2019. Modélisation multi-échelles de réseaux biologiques pour l'ingénierie métabolique d'un châssis biotechnologique. Ph.D. Thesis. Université Paris Saclay.

Tsakraklides, V., Kamineni, A., Consiglio, A.L., MacEwen, K., Friedlander, J., Blitzblau, H.G., Hamilton, M.A., Crabtree, D. V, Su, A., Afshar, J., Sullivan, J.E., LaTouf, W.G., South, C.R., Greenhagen, E.H., Shaw, A.J., Brevnova, E.E., 2018. High-oleate yeast oil without polyunsaturated fatty acids. Biotechnol. Biofuels 11:131. https://doi.org/10.1186/s13068-018-1131-y

Turck, D., Castenmiller, J., de Henauw, S., Hirsch-Ernst, K., Kearney, J., Maciuk, A., Mangelsdorf, I., McArdle, H.J., Naska, A., Pelaez, C., Pentieva, K., Siani, A.,

Thies, F., Tsabouri, S., Vinceti, M., Cubadda, F., Engel, K., Frenzel, T., Heinonen, M., Marchelli, R., Neuhäuser-Berthold, M., Pöting, A., Poulsen, M., Sanz, Y., Schlatter, J.R., van Loveren, H., Ackerl, R., Knutsen, H.K., 2019. Safety of *Yarrowia lipolytic*a yeast biomass as a novel food pursuant to regulation (EU) 2015/2283. EFSA J. 17. https://doi.org/10.2903/j.efsa.2019.5594

Turick, C.E., Knox, A.S., Becnel, J.M., Ekechukwu, A.A., Millike, C.E., 2010. Properties and Function of Pyomelanin, in: Biopolymers. Sciyo. https://doi.org/10.5772/10273

Vandermies, M., Denies, O., Nicaud, J.-M., Fickers, P., 2017. EYK1 encoding erythrulose kinase as a catabolic selectable marker for genome editing in the non-conventional yeast *Yarrowia lipolytica*. J. Microbiol. Methods 139, 161–164. https://doi.org/10.1016/j.mimet.2017.05.012

Verbeke, J., Beopoulos, A., Nicaud, J.-M., 2013. Efficient homologous recombination with short length flanking fragments in Ku70 deficient *Yarrowia lipolytica* strains. Biotechnol. Lett. 35, 571–576. https://doi.org/10.1007/s10529-012-1107-0

Vuralhan, Z., Luttik, M.A.H., Tai, S.L., Boer, V.M., Morais, M.A., Schipper, D., Almering, M.J.H., Kötter, P., Dickinson, J.R., Daran, J.M., Pronk, J.T., 2005. Physiological characterization of the ARO10-dependent, broad-substrate-specificity 2-oxo acid decarboxylase activity of *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 71, 3276–3284. https://doi.org/10.1128/AEM.71.6.3276-3284.2005

Vuralhan, Z., Morais, M.A., Tai, S.L., Piper, M.D.W., Pronk, J.T., 2003. Identification and characterization of phenylpyruvate decarboxylase genes in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 69, 4534–4541. https://doi.org/10.1128/AEM.69.8.4534-4541.2003

Wagner, J.M., Alper, H.S., 2016. Synthetic biology and molecular genetics in non-conventional yeasts: Current tools and future advances. Fungal Genet. Biol. FG B 89, 126–136. https://doi.org/10.1016/j.fgb.2015.12.001

Wagner, J.M., Williams, E. V, Alper, H.S., 2018. Developing a piggyBac Transposon System and Compatible Selection Markers for Insertional Mutagenesis and

Genome Engineering in *Yarrowia lipolytica*. Biotechnol. J. 13, e1800022. https://doi.org/10.1002/biot.201800022

Wang, H., Chen, H., Hao, G., Yang, B., Feng, Y., Wang, Y., Feng, L., Zhao, J., Song, Y., Zhang, H., Chen, Y.Q., Wang, L., Chen, W., 2013. Role of the phenylalanine-hydroxylating system in aromatic substance degradation and lipid metabolism in the oleaginous fungus *Mortierella alpina*. Appl. Environ. Microbiol. 79, 3225–3233. https://doi.org/10.1128/AEM.00238-13

Wang, H., Dong, Q., Meng, C., Shi, X. ai, Guo, Y., 2011. A continuous and adsorptive bioprocess for efficient production of the natural aroma chemical 2-phenylethanol with yeast. Enzyme Microb. Technol. 48, 404–407. https://doi.org/10.1016/J.ENZMICTEC.2011.01.006

Wang, Y., Chen, S., Yu, O., 2011. Metabolic engineering of flavonoids in plants and microorganisms. Appl. Microbiol. Biotechnol. 91, 949-956 https://doi.org/10.1007/s00253-011-3449-2

Wasylenko, T.M., Ahn, W.S., Stephanopoulos, G., 2015. The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in *Yarrowia lipolytica*. Metab. Eng. 30, 27–39. https://doi.org/10.1016/j.ymben.2015.02.007

Weber, W., Fussenegger, M., 2012. Emerging biomedical applications of synthetic biology. Nat. Rev. Genet. 13, 21-35. https://doi.org/10.1038/nrg3094

Wojtatowicz, M., Rymowicz, W., 1991. Effect of inoculum on kinetics and yield of citric acid production on glucose media by *Yarrowia lipolytica* A-101. Acta Aliment. Pol.

Wong, L., Engel, J., Jin, E., Holdridge, B., Xu, P., 2017. YaliBricks, a versatile genetic toolkit for streamlined and rapid pathway engineering in *Yarrowia lipolytica*. Metab. Eng. Commun. 5, 68–77. https://doi.org/10.1016/j.meteno.2017.09.001

Wu, M.R., Jusiak, B., Lu, T.K., 2019. Engineering advanced cancer therapies with synthetic biology. Nat. Rev. Cancer. 19, 187–195 https://doi.org/10.1038/s41568-019-0121-0

Xie, D., Jackson, E.N., Zhu, Q., 2015. Sustainable source of omega-3 eicosapentaenoic

acid from metabolically engineered *Yarrowia lipolytica*: from fundamental research to commercial production. Appl. Microbiol. Biotechnol. 99, 1599–1610. https://doi.org/10.1007/s00253-014-6318-y

Xu, P., Bhan, N., Koffas, M.A.G., 2013. Engineering plant metabolism into microbes: From systems biology to synthetic biology. Curr. Opin. Biotechnol. 24, 291–299 https://doi.org/10.1016/j.copbio.2012.08.010

Xue, Z., Sharpe, P.L., Hong, S.-P., Yadav, N.S., Xie, D., Short, D.R., Damude, H.G., Rupert, R.A., Seip, J.E., Wang, J., Pollak, D.W., Bostick, M.W., Bosak, M.D., Macool, D.J., Hollerbach, D.H., Zhang, H., Arcilla, D.M., Bledsoe, S.A., Croker, K., McCord, E.F., Tyreus, B.D., Jackson, E.N., Zhu, Q., 2013. Production of omega-3 eicosapentaenoic acid by metabolic engineering of *Yarrowia lipolytica*. Nat. Biotechnol. 31, 734–740. https://doi.org/10.1038/nbt.2622

Yang, W., Ruan, L., Tao, J., Peng, D., Zheng, J., Sun, M., 2018. Single Amino Acid Substitution in Homogentisate Dioxygenase Affects Melanin Production in *Bacillus thuringiensis*. Front. Microbiol. 9, 2242. https://doi.org/10.3389/fmicb.2018.02242

Yang, Z., Edwards, H., Xu, P., 2019. CRISPR-Cas12a/Cpf1-assisted precise, efficient and multiplexed genome-editing in *Yarrowia lipolytica*. Metab. Eng. Commun. e00112. https://doi.org/10.1016/j.mec.2019.e00112

Ye, M., Wang, Y., Qian, M., Chen, X., Hu, X., 2011. Preparation and Properties of the Melanin from *Lachnum singerianum*. International Journal of Basic & Applied Sciences IJBAS-IJENS. 11(03), 51-58

Young, E., Alper, H., 2010. Synthetic biology: Tools to design, build, and optimize cellular processes. J. Biomed. Biotechnol. https://doi.org/10.1155/2010/130781

Yovkova, V., Otto, C., Aurich, A., Mauersberger, S., Barth, G., 2014. Engineering the α-ketoglutarate overproduction from raw glycerol by overexpression of the genes encoding NADP+-dependent isocitrate dehydrogenase and pyruvate carboxylase in *Yarrowia lipolytica*. Appl. Microbiol. Biotechnol. 98, 2003–2013. https://doi.org/10.1007/s00253-013-5369-9

Zang, Y., Zha, J., Wu, X., Zheng, Z., Ouyang, J., Koffas, M.A.G., 2019. In Vitro

Naringenin Biosynthesis from p-Coumaric Acid Using Recombinant Enzymes. J. Agric. Food Chem. 67, 13430–13436. https://doi.org/10.1021/acs.jafc.9b00413

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Van Der Oost, J., Regev, A., Koonin, E. V., Zhang, F., 2015. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell 163, 759–771. https://doi.org/10.1016/j.cell.2015.09.038

Zhang, B., Chen, H., Li, M., Gu, Z., Song, Y., Ratledge, C., Chen, Y.Q., Zhang, H., Chen, W., 2013. Genetic engineering of *Yarrowia lipolytica* for enhanced production of trans-10, cis-12 conjugated linoleic acid. Microb. Cell Fact. 12, 70. https://doi.org/10.1186/1475-2859-12-70

Zhang, B., Rong, C., Chen, H., Song, Y., Zhang, H., Chen, W., 2012. De novo synthesis of trans-10, cis-12 conjugated linoleic acid in oleaginous yeast *Yarrowia lipolytica*. Microb. Cell Fact. 11, 51. https://doi.org/10.1186/1475-2859-11-51

Zhang, J. lai, Peng, Y.Z., Liu, D., Liu, H., Cao, Y.X., Li, B.Z., Li, C., Yuan, Y.J., 2018. Gene repression via multiplex gRNA strategy in *Y. lipolytica*. Microb. Cell Fact. 17:62. https://doi.org/10.1186/s12934-018-0909-8

6 Annexes

6.1 Supplementary data

According to the order of appearance in the main text, supplementary data from the articles presented during this manuscript are available hereafter. After which, supplementary information that were not developed on the presented articles is available (Codon optimized sequences of heterologous genes, table of *E. coli* and *Y. lipolytica* strains and table of main primers used).

6.1.1 Supplementary data of articles presented

The order in which the articles are presented here correspond to the order of appearance in the main text.

A modular Golden Gate toolkit for *Y. lipolytica* synthetic biology

Supplementary Table 1. List of strains and plasmids

Strains	*Genotype/plasmid*	references
Y. lipolytica		
W29	*Wild type*	(Barth and Gaillardin, 1996)
JMY195, Po1d	*MATA ura3-302 leu2-270 xpr2-322*	(Barth and Gaillardin, 1996)
JMY1212	Po1d *Δlip2 Δlip7 Δlip8-LEU2*-ZETA	(Emond, et al., 2010)
JMY5219	JMY1212 lys5::URA3ex	unpublished
JMY1350	JMY1212-URA3	This study
JMY7621	JMY1212 pTEF-RedstarII-Tlip-*URA3* ex	(Park, et al., 2018)
JMY7622	JMY1212 pGAP-RedstarII-Tlip-*URA3* ex	This study
JMY7623	JMY1212 pPGM-RedstarII-Tlip-*URA3* ex	This study
JMY7624	JMY1212 pTEF-2UAS-RedstarII-Tlip-*URA3* ex	This study
JMY7625	JMY1212 pTEF-4UAS-RedstarII-Tlip-*URA3* ex	This study
JMY7626	JMY1212 pTEF-8UAS-RedstarII-Tlip-*URA3* ex	This study
JMY7382	JMY1212 pEYK1-RedstarII-Tlip-*URA3* ex	(Park, et al., 2018)
JMY7627	JMY1212 pEYK1-2AB-RedstarII-Tlip-*URA3* ex	(Park, et al., 2018)
JMY7345	JMY1212 pEYK1-3AB-RedstarII-Tlip-*URA3* ex	(Park, et al., 2018)
JMY7628	JMY1212 pEYK1-4AB-RedstarII-Tlip-*URA3* ex	(Park, et al., 2018)
JMY7390	JMY1212 pEYK1-5AB-RedstarII-Tlip-*URA3* ex	(Park, et al., 2018)
JMY7253	JMY1212 pTEF-Turquoise-Tlip-*URA3* ex	This study
JMY7652	JMY1212 pTEF-YFP-Tlip-*URA3* ex	This study
JMY7653	JMY1212 pTEF-RedStar-TLip-pTEF-YFP-TLip-pTEF-Turquoise-Tlip-*URA3* ex	This study
JMY7655	Po1d pTEF-RedstarII-Tlip-*LEU2*	This study
JMY7656	JMY5219 pTEF-RedstarII-Tlip-*Lys5*	This study
JMY7657	JMY1212 pTEF-RedstarII-Tlip-*hph*	This study
JMY7658	JMY1212 pTEF-RedstarII-Tlip-*nat*	This study
JMY7659	JMY1212 pTEF-RedstarII-TTef-pEYK2AB-YFP-Tlip-*URA3* ex	This study
JMY7660	JMY1212 pTEF-RedstarII-TXpr2- pEYK2AB-YFP-Tlip-*URA3* ex	This study
JMY7661	JMY1212 pTEF-RedstarII-TGuo- pEYK2AB-YFP-Tlip-*URA3* ex	This study
JMY7662	JMY1212 pTEF-RedstarII-TSynth8- pEYK2AB-YFP-Tlip-*URA3* ex	This study
JMY7663	JMY1212 pTEF-RedstarII-Tlip- pEYK2AB-YFP-Tlip-*URA3* ex	This study

JMY5738	XYL+ Obese strain.	(Ledesma-Amaro, et al., 2016)	
JMY7767	Y1212 + GGA Xyl-Ura	This study	

E. coli	*Genotype/plasmid*	references	Addgene ID number
DH5**α**	Φ80*lacZ*Δm15 Δ(*lacZYA-arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (r$_k^-$, m$_k^+$) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ$^-$	(Promega)	
GGE029	pSB1A3-GB3	(Celinska, et al., 2017)	120730
GGE114	pSB1A3-ZetaUP-URA3-RFP-ZetaDOWN	(Celinska, et al., 2017)	120731
GGE083	pCR4Blunt-TOPO-M-URA3	(Celinska, et al., 2017)	120732
GGE176	pCR4Blunt-TOPO-M-LYS5	(Celinska, et al., 2017)	120733
GGE105	pCR4Blunt-TOPO-M-LEU2	This work	120734
GGE142	pCR4Blunt-TOPO-M-Suc2	This work	120735
GGE367	pCR4Blunt-TOPO-M-hph	This work	120736
GGE368	pCR4Blunt-TOPO-M-Nat	This work	120737
GGE085	pCR4Blunt-TOPO-P1 pTEF	(Celinska, et al., 2017)	120738
GGE002	pCR4Blunt-TOPO-P1 pGAPdh	(Celinska, et al., 2017)	120739
GGE001	pCR4Blunt-TOPO- P1 pPGM	(Celinska, et al., 2017)	120740
GGE145	pCR4Blunt-TOPO- P1 TEF-2UAS	(Celinska, et al., 2017)	120741
GGE146	pCR4Blunt-TOPO- P1 TEF-4UAS	(Celinska, et al., 2017)	120742
GGE147	pCR4Blunt-TOPO- P1 TEF-8UAS	(Celinska, et al., 2017)	120743
GGE0104	pUC57 - P1 pEYK1-3AB	(Park, et al., 2018)	120744
GGE0132	pUC57 - P1 pEYK1-4AB	(Park, et al., 2018)	120745
GGE250	pUC57 - P1 pEYK1-5AB	(Park, et al., 2018)	120746
GGE006	pCR4Blunt-TOPO-P2 pTEF	(Celinska, et al., 2017)	120747
GGE005	pCR4Blunt-TOPO-P2 pGAPdh	(Celinska, et al., 2017)	120748
GGE004	pCR4Blunt-TOPO- P2 pPGM	(Celinska, et al., 2017)	120749
GGE150	pCR4Blunt-TOPO- P2 TEF-2UAS	(Celinska, et al., 2017)	120750
GGE151	pCR4Blunt-TOPO- P2 TEF-4UAS	(Celinska, et al., 2017)	120751
GGE152	pCR4Blunt-TOPO- P2 TEF-8UAS	(Celinska, et al., 2017)	120752
GGE108	pCR4Blunt-TOPO- P2 pEYK1-3AB	This work	120753
GGE139	pCR4Blunt-TOPO- P2 pEYK1-4AB	This work	120754
GGE129	pCR4Blunt-TOPO- P2 pEYK1-5AB	This work	120755
GGE009	pCR4Blunt-TOPO-P3 pTEF	(Celinska, et al., 2017)	120756
GGE008	pCR4Blunt-TOPO-P3 pGAPdh	(Celinska, et al., 2017)	120757
GGE007	pCR4Blunt-TOPO-P3 pPGM	(Celinska, et al., 2017)	120758

GGE292	pCR4Blunt-TOPO-P3 TEF-2UAS	(Celinska, et al., 2017)	120759
GGE294	pCR4Blunt-TOPO-P3 TEF-4UAS	(Celinska, et al., 2017)	120760
GGE317	pCR4Blunt-TOPO-P3 TEF-8UAS	(Celinska, et al., 2017)	120761
GGE273	pCR4Blunt-TOPO-P3 pEYK1-3AB	This work	120762
GGE275	pCR4Blunt-TOPO-P3 pEYK1-4AB	This work	120763
GGE259	pCR4Blunt-TOPO-P3 pEYK1-5AB	This work	120764
GGE014	pCR4Blunt-TOPO-T1 Lip2	(Celinska, et al., 2017)	120765
GGE177	pCR4Blunt-TOPO-T1 XPR2	(Celinska, et al., 2017)	120766
GGE082	pCR4Blunt-TOPO-T1 TEF	(Celinska, et al., 2017)	120767
GGE025	pCR4Blunt-TOPO-T1 SynthGuo	(Celinska, et al., 2017)	120768
GGE027	pCR4Blunt-TOPO-T1 Synth8	(Celinska, et al., 2017)	120769
GGE015	pCR4Blunt-TOPO-T2 Lip2	(Celinska, et al., 2017)	120770
GGE257	pCR4Blunt-TOPO-T2 XPR2	This work	120771
GGE256	pCR4Blunt-TOPO-T2 TEF	This work	120772
GGE260	pCR4Blunt-TOPO-T2 SynthGuo	This work	120773
GGE080	pCR4Blunt-TOPO-T3 Lip2	(Celinska, et al., 2017)	120774
GGE258	pCR4Blunt-TOPO-T3 XPR2	This work	120775
GGE286	pCR4Blunt-TOPO-T3 TEF	This work	120776
GGE020	pCR4Blunt-TOPO-TLip2 (E-L)	(Celinska, et al., 2017)	120777
GGE 021	pCR4Blunt-TOPO-TLip2 (H-L)	(Celinska, et al., 2017)	120778
GGE077	pCR4Blunt-TOPO-G1-RedStarII	This work	120779
GGE270	pCR4Blunt-TOPO-G1-YFP	This work	120780
GGE190	pUC-G1-mTurquoise	This work (Twist)	120781
GGE070	pCR4Blunt-TOPO-G2-YFP	(Celinska, et al., 2017)	120782
GGE261	pCR4Blunt-TOPO-G3-mTurquoise	This work	120783
GGE067	pCR4Blunt-TOPO-Zeta-NotI_Up	(Celinska, et al., 2017)	120784
GGE038	pCR4Blunt-TOPO-Zeta-NotI_Down	(Celinska, et al., 2017)	120785
GGE091	pCR4Blunt-TOPO-Zeta-SfiI_Up	(Celinska, et al., 2017)	120786
GGE094	pCR4Blunt-TOPO-Zeta-SfiI_Down	(Celinska, et al., 2017)	120787
GGE255	pCR4Blunt-TOPO-Lip2-NotI_Up	This work	120788
GGE254	pCR4Blunt-TOPO-Lip2-NotI_Down	This work	120789
GGE253	pCR4Blunt-TOPO-Gsy-NotI_Up	This work	120790
GGE252	pCR4Blunt-TOPO-Gsy-NotI_Down	This work	120791
GGE207	pCR4Blunt-TOPO-MFE-NotI_Up	This work	120792

| GGE209 | pCR4Blunt-TOPO-MFE-NotI_Down | This work | 120793 |

A set of Y. lipolytica CRISPR/Cas9 vectors for exploiting wild type strains

supplementary data

Supplementary Table S1. *E. coli* and *Y. lipolytica* strains and plasmids used in this study

Strain (host strain)	Plasmid, genotype	References
***E. coli* strains**		
DH5**α**	Φ80*lacZ*Δm15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (r_k^-, m_k^+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ⁻	(Promega)
GGE029	GGE pSB1A3-GB3	(Celinska et al. 2017)
JME4315	pGM_sgRNA platform	This work
JME4473	p8UASTEF_CRISPR-Cas9_URA3ex_gGSY, AmpR	This work
JME4392	p8UASTEF_CRISPR-Cas9_LEU2ex_gGSY, AmpR	This work
JME4425	p8UASTEF_CRISPR-Cas9_LYS5ex_gGSY, AmpR	This work
JME4600	p8UASTEF_CRISPR-Cas9_NATex_gGSY, AmpR	This work
JME4759	p8UASTEF_CRISPR-Cas9_HPHex_gGSY, AmpR	This work
JME5019	p8UASTEF_CRISPR-Cas9_EYKex_gGSY, AmpR	This work
JME4452	p8UASTEF_CRISPR-Cas9_LEU2ex_gURA, AmpR	This work
JME4453	p8UASTEF_CRISPR-Cas9_LYS5ex_gURA, AmpR	This work
***Y. lipolytica* strains**		
W29	*Wild type*	(Barth and Gaillardin 1996)
Po1d, JMY195	*MATA leu2-270 ura3-302 xpr2-322*	(Barth and Gaillardin 1996)
JMY330	*MATA leu2-270 xpr2-322*	(Haddouche et al. 2010)
JMY2033	Po1d, *ura3::LEU2ex-zeta*	(Lazar et al. 2013)
JMY2394	Po1d *Δku70*	(Verbeke *et al.* 2013)
JMY5211	JMY2033 *lys5::URA3ex*	Unpublished
JMY5217	Po1d *lys5::URA3ex*	Unpublished
JMY7123	Po1d, Δ*lip2* Δ*lip7* Δ*lip8* ura3::LEU2-ZETA *lys5::URA3*ex Δ*eyk1*	Soudier *et al.* to be published
JMY7126	Po1d, Δ*lip2* Δ*lip7* Δ*lip8* ura3::*LEU2*-ZETA Δ*lys5* Δ*eyk1*	Soudier *et al.* to be published
JMY7923	*Wild-type* CBS 2074, Isolated from olives.	

JMY7926	*Wild-type* CBS 6125. Isolated from maize-processing plant.	
JMY7936	*Wild-type* CLIB 791. Isolated from goat's cheese	
JMY7937	*Wild-type* CLIB 879. Isolated from cheese.	
JMY7941	*Wild-type* DBVPG 4400. Isolated from commercial compost.	
JMY7942	*Wild-type* DBVPG 5851. Isolated from soil of palm tree.	
JMY7945	*Wild-type* NCYC 3271. Isolated from sugar cane peels.	
JMY7950	*Wild-type* PYCC 4743. Isolated from see water.	
JMY7181	*Wild-type.* IMUFRJ 50682 Isolated from estuary.	(Nunes et al. 2013)

Supplementary Table S2. Primers list

Primer name	Sequence	Use
GGP_E.coliVector_N_Fw	gcGGTCTCtGTCTtcttcctgcagtccggcaaaaaag	Amplification of *E. coli* part for constructing Cr Vector
GGP_E.coliVector_O_Rv	gcGGTCTCtTCAGaggctcttcgaattccagaaatcatc	Amplification of *E. coli* part for constructing Cr Vector
RFP-SfiI-Fw	ccGGCCATCTGGGCCctagagcaatacgcaaaccgc	Amplification of RFP for cloning into gRNA Platform
RFP_SfiI-Rv	ccGGCCCAGATGGCCtatataaacgcagaaaggcccac	Amplification of RFP for cloning into gRNA Platform
GSY_sgRNA_Fw	gggCGTCTCGtTTCGATTCCGGGTCGGCGCAGGTTGGCTGTTCGAGGTCGCCACCG	sgGSY cloning
GSY_sgRNA_Rv	cccCGTCTCtAGCTCTAAAACCGGTGGCGACCTCGAACAGCCAACCTGCGC	sgGSY cloning
T7_GSY_Fw	TATGGGTACAGGACGTGCATAG	GSY screening
T7_GSY_Rv	AGAGTAGCGTGTGTGGTGAAGA	GSY screening
URA_sgRNA_Fw	TTCGATTCCGGGTCGGCGCAGGTTGgAAGGAACTTGCTCTTAAGCAGTTTTA	sgURA cloning
URA_sgRNA_Rv	GCTCTAAAACTGCTTAAGAGCAAGTTCCTTcCAACCTGCGCCGACCCGGAAT	sgURA cloning
T7_URAorf_Fw	ATGCCCTCCTACGAAGCTCGAG	URA screening
T7_URAorf_Rv	GTTCTGGCCGTACAGACCTC	URA screening
Lip2_sgRNA_Fw	TTCGATTCCGGGTCGGCGCAGGTTGgATCTTCAAGCCCTTCAACTGGTTTTA	sgLip2 cloning
Lip2_sgRNA_Rv	GCTCTAAAACCAGTTGAAGGGCTTGAAGATcCAACCTGCGCCGACCCGGAAT	sgLip2 cloning
T7_Lip_Fw	CTTTCCACCATCCTCTTCACAG	Lip2 screening
T7_Lip_Rv	CTCCTGGCCAAAGAAGAGTTTA	Lip2 screening
gRNA1-MFE2-Fwd	TTCGATTCCGGGTCGGCGCAGGTTGGGACAAGGTTACCCGAGCTGCCGTTTTA	sgMFE2 cloning
gRNA1-MFE2-Rev	GCTCTAAAACGGCAGCTCGGGTAACCTTGTCCAACCTGCGCCGACCCGGAAT	sgMFE2 cloning

ForT7_MFE _Endon-Test-Fw-MFE	GTCGACGAGATTGTTTCCAAG	MFE2 screening
ForT7_MFE_Endon-Test-Rv-MFE	GGTCGCTTAACAGACTCCTCAG	MFE2 screening
gRNA-EYK1-Fwd	TTCGATTCCGGGTCGGCGCAGGTTGGCAAGCTCGTGCCCTCCGACAGTTTTA	sgEYK1 cloning
gRNA-EYK1-Rev	GCTCTAAAACTGTCGGAGGGCACGAGCTTGCAACCTGCGCCGACCCGGAAT	sgEYK1 cloning
T7_ EYK1_Fw	ATCTGTTCAACGAGACTGACGA	EYK1 screening
T7_EYK1_Rv	CCTGAACTTTACGAAGGATCG	EYK1 screening
ht17-gRNA-EYD1-Fwd	TTCGATTCCGGGTCGGCGCAGGTTGGTCAGTCTGCCTCGCCGAGCTGTTTTA	sgEYD1 cloning
ht18-gRNA-EYD1-Rev	GCTCTAAAACAGCTCGGCGAGGCAGACTGACCAACCTGCGCCAACCCGGAAT	sgEYD1 cloning
ht35-Verif-EYD1-Fwd	GCCCATCTCGGCACCCTACA	EYD1 screening
ht36-Verif-EYD1-Rev	GGCTCAAACTGGCCGAGCTC	EYD1 screening

Barth G, Gaillardin C (1996) Yarrowia lipolytica. In: Wolf K, Breuning KD, Barth J (eds) Non-conventional Yeasts in Biotechnology. Springer-Verlag, Berlin, pp 313-388

Celinska E, Ledesma-Amaro R, Larroude M, Rossignol T, Pauthenier C, Nicaud JM (2017) Golden Gate Assembly system dedicated to complex pathway manipulation in Yarrowia lipolytica Microb Biotechnol 10:450-455 doi:10.1111/1751-7915.12605

Haddouche R, Delessert S, Sabirova J, Neuvéglise C, Poirier Y, Nicaud J-M (2010) Roles of multiple acyl-CoA oxidases in the routing of carbon flow towards β-oxidation and polyhydroxyalkanoate biosynthesis in *Yarrowia lipolytica* FEMS Yeast Res 10:917-927 doi:10.1111/j.1567-1364.2010.00670.x

Lazar Z, Rossignol T, Verbeke J, Crutz-Le Coq A-M, Nicaud J-M, Robak M (2013) Optimized invertase expression and secretion cassette for improving *Yarrowia lipolytica* growth on sucrose for industrial applications J Ind Microbiol Biotechnol 40:1273-1283 doi:10.1007/s10295-013-1323-1

Nunes PMB, da Rocha SM, Amaral PFF, da Rocha-Leão MHM (2013) Study of trans–trans farnesol effect on hyphae formation by Yarrowia lipolytica Bioprocess Biosystems Eng 36:1967-1975 doi:10.1007/s00449-013-0973-8

Verbeke J, Beopoulos A, Nicaud J-M (2013) Efficient homologous recombination with short length flanking fragments in Ku70 deficient Yarrowia lipolytica strains Biotechnol Lett 35:571-576 doi:10.1007/s10529-012-1107-0

6.1.2 **Codon optimized sequences of heterologous genes**

Underlined are highlighted the GG recognition sites followed by their 4-nt over-hang. In Red are highlighted BamHI recognition sites and in blue AvrII sites, both used for cloning genes into JMP62 vectors. In yellow are highlighted the CYC-t terminators and in green the SV40 NLS used in Cas9 and Cpf1 enzymes.

| SpCas9 | GGGGGTCTCtACAAccacaATGGATAAGAAATACTCCATTGGCCTGGA-
CATCGGAACCAACTCCGTGGGTTGGGCCGTGATCACCGATGAGTACAAGGTGCCCTCTAAGAAATTCAAGGTCCTGGGC
AACACCGACCGACACTCCATCAAGAAGAACCTGATCGGCGCTCTGCTCTTCGACTCTGGCGA-
GACTGCTGAGGCCACCCGACTGAAGCGAACCGCTCGAAGACGATACACCCGAAGAAGAACCGAATCTGTTACCTGCAG
GAGATCTTCTCTAACGAGATGGCCAAGGTGGACGACTCTTTCTTCCACCGACTGGAG-
GAGTCTTTCCTGGTGGAGGAGGACAAGAAGCACGAGCGACACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACC
ACGAGAAGTACCCCACCATCTACCACCTGCGAAAGAAGCTGGTGGACTCTACCGA-
CAAGGCCGACCTGCGACTGATCTACCTGGCCCTGGCCCACATGATCAAGTTCCGAGGCCACTTCCTGATCGAGGGCGAC
CTGAACCCCGACAACTCTGACGTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTA-
CAACCAGCTCTTCGAAGAGAACCCCATTAACGCTTCTGGCGTGGATGCTAAGGCCATCCTGTCTGCCCGACTGTCTAAG
TCTCGACGACTCGAGAACCTGATTGCTCAGCTCCCCGGAGAGAAGAAGAACGG-
TCTGTTCGGAAACCTGATTGCTCTGTCCCTGGGTCTGACCCCTAACTTCAAGTCCAACTTCGATCTGGCTGAGGACGCT
AAGCTGCAGCTGTCTAAGGACACCTACGACGATGACCTGGATAACCTGCTCGCCCAGATT-
GGCGACCAGTACGCCGACCTGTTCCTGGCCGCCAAGAACCTGTCTGACGCCATCCTGCTGTCTGACATCCTGCGAGTGA
ACACCGAGATCACCAAGGCCCCCCTGTCTGCCTCCATGATTAAGCGATACGATGAGCACCAC-
CAGGATCTGACCCTCCTCAAGGCTCTGGTCCGACAGCAGCTGCCCGAGAAGTACAAGGAGATTTTCTTCGACCAGTCTA
AGAACGGCTACGCCGGCTACATCGACGGCGGCGCCTCTCAGGAGGAGTTCTACAAGTTCAT-
TAAGCCCATCCTGGAGAAGATGGACGGAACCGAGGAACTGCTCGTGAAGCTGAACCGAGAGGACCTCCTGCGAAAGCAG
CGAACCTTCGACAACGGCTCTATCCCCCACCA-
GATCCACCTGGGCGAGCTGCACGCCATCCTGCGACGACAGGAGGACTTCTACCCCTTCCTGAAGGACAACCGAGAGAAG
ATCGAGAAGATCCTGACCTTCCGAATCCCCTACTACGTGGGaCCCCTGGCCCGAG-
GaAACTCTCGATTCGCtTGGATGACCCGAAAGTCTGAGGAGACTATTACCCCCTGGAACTTCGAGGAGGTGGTGGATAA
GGGCGCCTCTGCTCAGTCTTTCATCGAGCGAATGACCAACTTCGA-
CAAGAACCTCCCCAACGAGAAGGTCCTGCCCAAGCACTCTCTGCTCTACGAGTACTTCACCGTCTACAACGAGCTCACC
AAGGTCAAGTACGTGACCGAGGGAATGCGAAAGCCCGCTTTCCTGTCTGGAGAGCA-
GAAGAAGGCTATTGTGGATCTGCTCTTCAAGACTAACCGAAAGGTCACCGTCAAGCAGCTGAAGGAGGATTACTTCAAG
AAGATTGAGTGTTTCGATTCTGTCGAGATCTCCGGCGTCGAG-
GACCGATTCAACGCCTCTCTGGGTACCTACCACGACCTGCTGAAGATTATCAAGGACAAGGATTTCCTGGATAACGAGG
AGAACGAGGATATTCTCGAGGACATTGTCCTGACCCTCACCCTGTTCGAGGATCGAGAGAT-
GATTGAGGAGCGACTCAAGACCTACGCTCACCTGTTCGACGACAAGGTGATGAAGCAGCTGAAGCGACGACGATACACC
GGCTGGGGCCGACTGTCTCGAAAGCTGATCAACGGCATCCGAGA-
CAAGCAGTCTGGCAAGACCATCCTGGACTTCCTGAAGTCTGACGGCTTCGCCAACCGAAACTTCATGCAGCTGATCCAC
GACGACTCTCTGACCTTCAAGGAGGACATCCA-
GAAGGCCCAGGTGTCTGGCCAGGGCGACTCTCTGCACGAGCACATCGCCAACCTGGCCGGCTCTCCCGCCATTAAGAAA
GGTATCCTGCAGACCGTCAAGGTGGTCGATGAGCTCGTCAAGGTGATGGGCCGACA-
CAAGCCCGAGAACATTGTCATTGAGATGGCTCGAGAGAACCAGACTACTCAGAAGGGTCAGAAAAACTCCCGAGAGCGA
ATGAAGCGAATTGAGGAAGGTATTAAGGAGCTGGGCTCCCAGATTCTCAAG-
GAGCATCCCGTGGAGAACACTCAGCTCCAGAACGAGAAGCTGTACCTGTACTATCTGCAGAACGGTCGAGACATGTACG
TCGACCAGGAGCTGGATATCAACCGACTCTCCGACTACGATGTGGACCA-
CATTGTGCCCCAGTCCTTCCTGAAGGACGATTCTATCGATAACAAGGTGCTGACCCGATCCGACAAGAACCGAGGCAAG
TCTGACAACGTGCCCTCCGAGGAGGTGGTCAAGAAGATGAAGAACTACTGGCGACAGCTGCT-
GAACGCCAAGCTGATTACCCAGCGAAAGTTCGACAACCTGACCAAGGCCGAGCGAGGCGGCCTGTCTGAGCTGGACAAG
GCCGGCTTCATCAAGCGACAGCTGGTGGAGACTCGACAGATCACCAAGCACGTGGCCCA-
GATCCTGGACTCTCGAATGAACACCAAGTACGACGAGAACGACAAGCTGATTCGAGAGGTGAAGGTGATCACCCTGAAG
TCTAAGCTGGTGTCTGACTTCCGAAAGGACTTCCAGTTCTACAAGGTGCGAGA-
GATTAACAACTACCACCACGCCCACGATGCCTACCTGAACGCTGTCGTGGGCACCGCCCTCATCAAGAAGTATCCCAAG
CTGGAGTCCGAGTTCGTCTACGGCGACTACAAGGTCTACGATGTGCGAAAAATGATT-
GCCAAGTCCGAGCAGGAGATTGGCAAGGCTACCGCCAAGTACTTCTTCTACTCCAACATTATGAACTTCTTCAAGACCG
AGATTACCCTGGCTAACGGCGAGATTCGAAAGCGACCCCTCATTGAGACTAACGGAGA-
GACTGGTGAGATCGTGTGGGACAAGGGACGAGACTTCGCCACCGTGCGAAAGGTGCTGTCTATGCCCCAGGTGAACATC
GTGAAGAAGACCGAGGTGCAGACCGGAGGTTTCTCTAAG-
GAGTCCATCCTGCCCAAGCGAAACTCTGACAAGCTGATCGCCCGAAAGAAGGACTGGGACCCCAAGAAGTACGGAGGTT
TCGACTCTCCCACCGTGGCTTACTCTGTGCTGGTGGTGGCCAAGGTGGA-
GAAGGGCAAGTCTAAGAAGCTGAAGTCTGTGAAGGAGCTGCTGGGCATTACCATCATGGAGCGATCTTCTTTCGAGAAG
AACCCCATTGACTTCCTGGAGGCCAAGGGATACAAGGAGGTGAAGAAGATCTGAT-
TATCAAGCTCCCCAAGTACTCTCTGTTCGAGCTGGAGAACGGACGAAAGCGAATGCTGGCCTCTGCCGGCGAGCTGCAG
AAGGGAAACGAGCTGGCCCTGCCCTCCAAGTACGTCAACTTCCTGTACCTCGCCTCCCAT-
TACGAGAAGCTGAAGGGCTCTCCCGAGGATAACGAGCAGAAGCAGCTCTTCGTGGAGCAGCATAAGCACTACCTGGACG
AGATCATCGAGCAGATTTCTGAGTTCTCTAAGCGAGTGATCCTGGCT-
GACGCCAACCTGGATAAGGTGCTGTCTGCTTACAACAAGCACCGAGACAAGCCCATTCGAGAGCAGGCTGAGAACATCA
TTCACCTGTTCACCCTGACCAACCTGGGAGCCCCCGCTGCCTTCAAGTACTTCGACACCAC-
CATCGACCGAAAGCGATACACCTCTACCAAGGAGGTGCTGGACGCCACCCTGATCCACCAGTCTATCACCGGCCTGTAC
GA-
GACTCGAATCGACCTGTCTCAGCTGGGCGGCGACTCTCGAGCCGACCCCAAGAAGAAGCGAAAGGTGTAAGCTAGCCTC |

	ATGTA<mark>ATTAGTTATGTCACGCTTACATTCACGCCCTCCCCCCA-</mark> <mark>CATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGTCTAGGTCCCTATTTATTTTTTTATAGTTATGTTAGT</mark> <mark>ATTAAGAACGTTATTTATATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCGTG-</mark> <mark>TACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGTTTTGGGACGCTCGAAGGCTTTAATTTGCAAGCT</mark>TGGCG <u>GTCTa</u>GAGACCCCC
FnCpf1	GGGGG<u>TCTC</u>t<u>ACAA</u>ccaca<u>A</u>TGTCTATCTACCAGGAGTTCGTGAACAAG- TACTCTCTGTCTAAGACCCTGCGATTCGAGCTGATCCCCCAGGGCAAGACCCTGGAGAACATCAAGGCCCGAGGCCTGA TCCTGGACGACGAGAAGCGAGCCAAGGACTACAAGAAGGCCAAGCAGATCATCGACAAGTAC- CACCAGTTCTTCATCGAGGAGATCCTGTCTTCTGTGTGCATCTCTGAGGACCTGCTGCAGAACTACTCTGACGTGTACT TCAAGCTGAAGAAGTCTGACGACGACAACCTGCAGAAGGACTTCAAGTCTGCCAAGGACAC- CATCAAGAAGCAGATCTCTGAGTACATCAAGGACTCTGAGAAGTTCAAGAACCTGTTCAACCAGAACCTGATCGACGCC AAGAAGGGCCAGGAGTCTGACCTGATCCTGTGGCTGAAGCAGTCTAAGGA- CAACGGCATCGAGCTGTTCAAGGCCAACTCTGACATCACCGACATCGACGAGGCCCTGGAGATCATCAAGTCTTTCAAG GGCTGGACCACCTACTTCAAGGGCTTCCACGAGAACCGAAAGAACGTGTACTCTTCTAACGA- CATCCCCACCTCTATCATCTACCGAATCGTGGACGACAACCTGCCCAAGTTCCTGGAGAACAAGGCCAAGTACGAGTCT CTGAAGGACAAGGCCCCCGAGGCCATCAACTACGAGCAGATCAAGAAGGACCTGGCCGAG- GAGCTGACCTTCGACATCGACTACAAGACCTCTGAGGTGAACCAGCGAGTGTTCTCTCTGGACGAGGTGTTCGAGATCG CCAACTTCAACAACTACCTGAACCAGTCTGGCATCACCAAGTTCAACAC- CATCATCGGCGGCAAGTTCGTGAACGGCGAGAACACCAAGCGAAAGGGCATCAACGAGTACATCAACCTGTACTCTCAG CAGATCAACGACAAGACCCTGAAGAAGTACAAGATGTCTGTGCTGTTCAAGCA- GATCCTGTCTGACACCGAGTCTAAGTCTTTCGTGATCGACAAGCTGGAGGACGACTCTGACGTGGTGACCACCATGCAG TCTTTCTACGAGCAGATCGCCGCCTTCAAGACCGTGGAGGAGAAGTCTATCAAGGA- GACtCTGTCTCTGCTGTTCGACGACCTGAAGGCCCAGAAGCTGGACCTGTCTAAGATCTACTTCAAGAACGACAAGTCT CTGACCGACCTGTCTCAGCAGGTGTTCGACGACTACTCTGT- GATCGGCACCGCCGTGCTGGAGTACATCACCCAGCAGATCGCCCCCAAGAACCTGGACAACCCCTCTAAGAAGGAGCAG GAGCTGATCGCCAAGAAGACCGAGAAGGCCAAGTACCTGTCTCTGGA- GACtATCAAGCTGGCCCTGGAGGAGTTCAACAAGCACCGAGACATCGACAAGCAGTGTCGATTCGAGGAGATCCTGGCC AACTTCGCCGCCATCCCCATGATCTTCGACGAGATCGCCCAGAACAAGGACAACCTGGCCCA- GATtTCTATCAAGTACCAGAACCAGGGCAAGAAGGACCTGCTGCAGGCCTCTGCCGAGGACGACGTGAAGGCCATCAAG GACCTGCTGGACCAGACCAACAACCTGCTGCACAAGCTGAAGATCTTCCA- CATCTCTCAGTCTGAGGACAAGGCCAACATCCTGGACAAGGACGAGCACTTCTACCTGGTGTTCGAGGAGTGCTACTTC GAGCTGGCCAACATCGTGCCCCTGTACAACAAGATCCGAAACTACATCACCCA- GAAGCCCTACTCTGACGAGAAGTTCAAGCTGAACTTCGAGAACTCTACCCTGGCCAACGGCTGGGACAAGAACAAGGAG CCCGACAACACCGCCATCCTGTTCATCAAGGACGACAAGTACTACCTGGGCGTGAT- GAACAAGAAGAACAACAAGAT tTTCGACGACAAGGCCATCAAGGAGAACAAGGGCGAGGGCTACAAGAAGATCGTGTAC AAGCTGCTGCCCGGCGCCAACAAGATGCTGCCCAAGGTGTTCTTCTCTGCCAAGTC- TATCAAGTTCTACAACCCCTCTGAGGACATCCTGCGAATCCGAAACCACTCTACCCACCAAGAACGGCTCTCCCCAG AAGGGCTACGAGAAGTTCGAGTTCAACATCGAGGACTGCCGAAAGTTCATCGACTTCTA- CAAGCAGTCTATCTCTAAGCACCCCGAGTGGAAGGACTTCGGCTTCCGATTCTCTGACACCCAGCGATACAACTCTATC GACGAGTTCTACCGAGAGGTGGAGAACCAGGGCTACAAGCTGACCTTCGAGAACATCTCT- GAGTCTTACATCGACTCTGTGGTGAACCAGGGCAAGCTGTACCTGTTCCAGATCTACAACAAGGACTTCTCTGCCTACT CTAAGGGCCGACCCAACCTGCACACCCTG- TACTGGAAGGCCCTGTTCGACGAGCGAAACCTGCAGGACGTGGTGTACAAGCTGAACGGCGAGGCCGAGCTGTTCTACC GAAAGCAGTCTATCCCCAAGAAGATCACCCACCCCGCCAAG- GAGGCCATCGCCAACAAGAACAAGGACAACCCCAAGAAGGAGTCTGTGTTCGAGTACGACCTGATCAAGGACAAGCGAT TCACCGAGGACAAGTTCTTCTTCCACTGCCCCATCAC- CATCAACTTCAAGTCTTCTGGCGCCAACAAGTTCAACGACGAGATCAACCTGCTGCTGAAGGAGAAGGCCAACGACGTG CACATCCTGTCTATCGACCGAGGCGAGCGACACCTGGCCTACTACACCCTGG- TGGACGGCAAGGGCAACATCATCAAGCAGGACACCTTCAACATCATCGGCAACGACCGAATGAAGACCAACTACCACGA CAAGCTGGCCGCCATCGAGAAGGACCGAGACTCTGCCCGAAAG- GACTGGAAGAAGATCAACAACATCAAGGAGATGAAGGAGGGCTACCTGTCTCAGGTGGTGCACGAGATCGCCAAGCTGG TGATCGAGTACAACGCCATCGTGGTGTTCGAGGACCT- GAACTTCGGCTTCAAGCGAGGCCGATTCAAGGTGGAGAAGCAGGTGTACCAGAAGCTGGAGAAGATGCTGATCGAGAAG CTGAACTACCTGGTGTTCAAGGACAACGAGTTCGA- CAAGACCGGCGGCGTGCTGCGAGCCTACCAGCTGACCGCCCCCTTCGAGACtTTCAAGAAGATGGGCAAGCAGACCGGC ATCATCTACTACGTGCCCGCCGGCTTCACCTCTAAGATtTGCCCCGTGACCGGCTTCGT- GAACCAGCTGTACCCCAAGTACGAGTCTGTGTCTAAGTCTCAGGAGTTCTTCTCTAAGTTCGACAAGATCTGCTACAAC CTGGACAAGGGCTACTTCGAGTTCTCTTTTCGACTACAAGAACTTCGGCGA- CAAGGCCGCCAAGGGCAAGTGGACCATCGCCTCTTTCGGCTCTGCACTGATCAACTTCCGAAACTCTGACAAGAACCAC AACTGGGACACCCGAGAGGTGTACCCCACCAAGGAGCTGGAGAAGCTGCTGAAGGACTACTC- TATCGAGTACGGCCACGGCGAGTGCATCAAGGCCGCCATCTGCGGCGAGTCTGACAAGAAGTTCTTCGCCAAGCTGACC TCTGTGCTGAACACCATCCTGCAGATGCGAAACTCTAAGACCGGCACCGAGCTGGACTACCT- GATCTCTCCCGTGGCCGACGTGAACGGCAACTTCTTCGACTCTCGACAGGCCCCCAAGAACATGCCCCAGGACGCCGAC GCCAACGGCGCCTACCACATCGGCCTGAAGGGCCT- GATGCTGCTGGGCCGAATCAAGAACAACCAGGAGGGCAAGAAGCTGAACCTGGTGATCAAGAACGAGGAGTACTTCGAG TTCGTGCAGAACCGAAACAAC<mark class="green">CCCAAGAAGAAGCGAAAGGTG</mark>TAAGCTAGCCTCATGTA<mark>AT-</mark> <mark>TAGTTATGTCACGCTTACATTCACGCCCTCCCCCCACATCCGCTCTAACCGAAAAGGAAGGAGTTAGACAACCTGAAGT</mark> <mark>CTAGGTCCCTATTTATTTTTTTATAGTTATGTTAGTATTAAGAACGTTATTTA-</mark> <mark>TATTTCAAATTTTTCTTTTTTTTCTGTACAGACGCGTGTACGCATGTAACATTATACTGAAAACCTTGCTTGAGAAGGT</mark> <mark>TTTGGGACGCTCGAAGGCTTTAATTTGCAAGCT</mark>TGGCG<u>GTCTa</u>GAGAC<u>C</u>CCCC
Sc ARO4- K229L	GGGG<u>TCTC</u>TAAT<u>G</u>TCTGAGTCCCCTATGTTTGCAGCTAATGG- TATGCCTAAAGTCAACCAGGGAGCCGAGGAAGACGTGCGAATCTTGGGATATGATCCGTTGGCTTCCCCGGCCTTGCTG CAGGTTCAGATTCCCGCGACTCCCACCTCGCTCGAGACAGCAAAACGAGGCCGCCGA- GAGGCCATCGACATCATCACTGGCAAGGATGATCGAGTCTTGGTTATCGTGGGCCCCTGCTCGATACACGACCTGGAAG

_G1	CTGCCCAGGAGTATGCACTCCGACTTAAGAAGCTGAGCGACGAGCTCAAGGGCGACTT- GTCGATTATCATGCGGGCTTATCTGGAAAAGCCCCGAACTACTGTTGGATGGAAGGGCTTGATCAACGATCCCGACGTG AATAATACATTTAATATTAATAAAGGACTTCAGTCGGCCCGTCAGCTGTTCGTTAAT- TTAACCAACATTGGCCTGCCCATCGGATCAGAGATGCTGGCACAATCAGTCCCCAGTATCTTGCCGACCTCGTGTCTT TTGGCGCGATCGGAGCACGAACTACTGAGTCGCAGTTGCATCGAGAGCTCGCTTCGGGCT- TATCCTTTCCCGTGGGATTTAAGAATGGCACTGACGGAACTTTGAACGTCGCCGTTGACGCATGCCAGGCTGCCGCGCA CTCGCATCACTTTATGGGCGTCACCTTGCACGGAGTCGCGGCCATTACTA- CAACAAAAGGCAATGAGCATTGTTTTGTCATCTTGCGAGGCGGAAAGAAAGGAACTAATTATGATGCAAAATCTGTCGC CGAGGCAAAAGCCCAGTTACCGGCGGGCTCAAATGGATTGATGATCGATTA- TAGCCATGGAAATAGCAACAAAGACTTTCGAAATCAGCCGAAAGTGAACGATGTGGTCTGCGAACAGATAGCGAATGGC GAGAATGCAATCACAGGAGTGATGATCGAGAGTAATATTAAT- GAGGGCAATCAGGGAATTCCCGCGGAAGGCAAGGCTGGACTTAAGTACGGAGTCTCTATAACAGACGCGTGCATTGGCT GGGAAACCACCGAGGATGTTCTGCGAAAGCTGGCCGCCGCAGTGCGCCAGCGGCGT- GAGGTGAATAAGAAGTAATCTAAGAGACCCC
Sc ARO7-T226I _G2	GGGGGTCTCTACAATGGACTTTACTAAGCCTGAAACCGTGCTGAACTTGCAGAACA- TACGCGACGAGTTGGTCCGGATGGAAGACAGCATTATCTTTAAGTTCATCGAACGCTCTCACTTTGCTACTTGCCCATC GGTGTACGAAGCGAATCACCCTGGACTT- GAGATACCCAACTTCAAGGGCTCCTTTCTGGACTGGGCCCTGTCTAACCTGGAGATCGCTCACTCGCGGATTCGCCGAT TTGAGTCGCCCGACGAGACACCGTTCTTCCCCGATAAAATCCAAAGTCTTTCTT- GCCTTCCATAAATTATCCGCAGATACTCGCTCCCTACGCTCCTGAGGTGAACTATAACGACAAGATTAAGAAGGTGTAC ATCGAGAAAATCATCCCTCTCATCAGTAAGCGAGACGGAGACGA- CAAGAACAACTTTTCAGTGGCGACCCGAGACATTGAGTGCCTGCAGTCTCTGTCTCGTCGTATACATTTCGGAAAATTC GTCGCAGAGGCTAAATTTCAGTCAGACATTCCCTTGTATACCAAATTGATTAAGAGCAAG- GACGTGGAAGGTATCATGAAGAACATTACTAACTCCGCAGTCGAGGAGAAAATACTGGAGCGCTTGACAAAGAAAGCAG AGGTTTACGGAGTTGATCCCACAAATGAATCCGGAGAGCGTCGAATCATCCCTGAG- TACTTAGTCAAGATCTACAAAGAGATCGTGATTCCCATAACCAAAGAGGTCGAAGTCGAGTATCTTCTCCGTCGACTGG AAGAATAATCTAAGAGACC
Sc Aro3-K222L _G2	ggGGTCTCtACAAATGTTTATCAAGAATGACCATGCAGGCGATCGTAAGCGACTCGAGGATT- GGCGAATTAAGGGATACGACCCTTTGACTCCCCCTGACCTCTTGCAGCACGAGTTCCCTATCTCCGCTAAGGGCGAAGA GAATATCATTAAAGCCCGCGATAGTGTTTGCGACATCCTCAACGGCAAAGGACGACCGATTGG- TGATTGTTATTGGCCCCTGCTCGTTGCACGATCCGAAGGCTGCATATGACTATGCCGATCGATTAGCCAAGATCTCTGA GAAATTATCGAAGGATTTGTTAATCATAATGCGAGCATACTTGGA- GAAGCCTCGAACCACCGTGGGTTGGAAGGGCTTAATCAATGATCCCGACATGAACAATTCGTTCCAGATTAACAAGGGA TTGCGAATCTCTCGAGAGATGTTTATCAAGTTGGTCGAGAAGTT- GCCAATCGCCGGCGAAATGCTGGACACTATCAGCCCCCAATTCCTTTCCGACTGCTTTTCTCTGGGAGCTATTGGTGCT CGAACAACCGAGTCTCAGTTGCATCGTGAGCTGGCCAGCGGCCTGTCCTTTCCCA- TAGGTTTCAAGAATGGAACCGACGGCGGCCTGCCAGGTTGCTATTGATGCGATGCGAGCTGCCGCCCACGAGCACTATTT TCTCTCCGTGACCTTGCCGGGAGTGACCGCCATTGTTGGTACCGAGGGCAATAAAGACA- CATTCTTGATTCTCCGAGGCGGCAAGAATGGCACAAATTTCGACAAGGAGTCGGTGCAGAACACCAAGAAGCAATTGGA GAAAGCTGGCCTCACAGACGACAGTCAAAAGCGAATCATGATTGACTGCTCTCATGGAAA- TAGCAACAAGGACTTTAAGAATCAGCCCAAAGTGGCTAAGTGCATCTACGATCAATTGACCGAAGGCGAAAACAGCTTG TGCGGCGTCATGATCGAATCTAATATCAACGAGGGACGTCAGGA- CATCCCTAAAGAGGGCGGACGCGAAGGCCTCAAATACGGATGCTCGGTGACTGACGCCTGCATCGGATGGGAAAGCACT GAGCAAGTGCTCGAATTGCTCGCGGAAGGCGTCCGCAATCGACGAAAAGCACTTAAGAAGTAAGGATaGAGACCgg
AtPAL1 _G1	GGGGTCTCTAATGGAAATCAACGGAGCTCATAAATCTAATGGCGGCGG- TGTCGATGCAATGTTGTGTGGTGGCGATATTAAAACCAAGAATATGGTCATTAATGCCGAAGACCCGTTGAATTGGGGC GCAGCCGCTGAACAGATGAAGGGCTCTCACCTTGACGAGGTCAAACGAATGGTCGCA- GAATTCCGAAAACCTGTCGTCAACTTGGGCGGAGAAACCTTGACAATCGGCCAGGTCGCCGCCATTTCCACCATCGGAA ATTCCGTTAAAGTTGAACTCTCAGAAACTGCACGAGCGGGCGTTAACGCCTCGTCT- GACTGGGTGATGGAATCTATGAATAAGGGACAGACAGCTACGGCCGTGACCACCGGATTCGGAGCCACCTCGCACCGTC GCACTAAGAATGGCGTTGCTTTACAAAAGGGAGTTGATACGGTTTCTCAATGCTGGCATCTTT- GGTTCCACCAAGGAGACTTCGCATACCCTGCCCCATTCTGCTACCCGAGCTGCTATGTTGGTCCGCATTAATACCCTGT TGCAGGGCTTCTCTGGAATAAGATTCGAAATCTTGGAAGCTATCACTAGCTTCTTGAATAA- TAATATTACACCTAGCCTGCCGTTGCGCGGAACCATTACTGCTAGCGGTGACTTGGTGCCATTGTCATATATTGCAGGT TTGTTGACTGGACGGCCGAACTCAAAGGCCACCGGACCTAATGGAGAGGCCTT- GACTGCGGAAGAGGCATTTAAGTTGGCCGGTATTTCCTCTGGTTTCTTCGACCTTCAACCAAAAGAGGGATTGGCTTTG GTTAACGGTACTGCAGTCGGTTCGGGTATGGCCAGTATGGTTCTCTTTGAAACCAACGTGTT- GAGCGTGTTAGCAGAAATCCGTCTCCGCCGTGTGTTGCCGAAGTCATCTGGGGAAAACCCGAATTTACTGACCACTTGACC CATCGTTTGAAGCACCACCTGGACAGATTGAGGCCGCAGCCATTATGGAACACATTTT- GGATGGCTCTTCATATATGAAGCTCGCGCAAAAGTTGCATGAAATGGACCCCTTGCAAAAGCCCAAGCAGGACCGATAT GCCTTGCGCAACCAGTCCCCAGTGGCTGGGACCCCAGATTGAGGTTATTCGCTATCCA- CAAAGTCCATTGAAAGAGAAATCAATTCGGTTAATGACAACCCACTGATTGACGTCTCTCGAAATAAAGTATCCATGG CGGAAATTTTCAGGGTACTCCTATAGGTGTGAGCATGGACAA- TACCCGGCTCGCCATCGCCGCCATCGGCAAGTTGATGTTCGCCCAGTTTAGTGAATTGGTCAACGACTTTTATAATAAC GGACTCCCTAGTAACTTGACTGCCTCTCGCAACCCTTCCTTA- GACTACGGCTTTAAAGGCGCCGAAATAGCTATGGCATCGTACTGCTCGGAATTGCAGTATTTGGCCAACCCAGTCACCT CGCACGTCCAGTCCGCTGAACAGCACAATCAGGACGTTAATTCACTCGGCCT- GATTTCATCGCGGAAGACATCTGAGGCCGTCGACATCTTGAAATTGATGTCCACTACATTCTTAGTCGCCATCTGCCAG GCAGTCGACCTTCGGCACCTCGAAGAAAACTTACGGCAAACCGTTAAGAA- TACAGTGTCGCAGGTTGCAAAGAAGGTGTTGACAACCGGTGTGAACGGAGAATTGCACCCCAGTCGATTTTGTGAGAAA GACTTGTTGAAGGTGGTCGATCGAGAGCAGGTTTATACCTATGCA- GACGACCCCTGCTCCGCTACTTATCCACTCATCCAGAAATTGCGGCAGGTCATAGTGGATCACGCCCTCATTAACGGAG

	AATCTGAAAAGAACGCCGTTACCTCTATCTTTCACAAAATCGGTGCCTTT- GAAGAAGAATTGAAAGCCGTTTTGCCCAAGGAAGTCGAGGCTGCCCGAGCCGCCGCTTATGACAATGGTACCAGTGCAATTC CCAATCGAATTAAAGAGTGCCGAAGTTACCCTCTGTACCGATTTGTTCGAGAGGAAT- TAGGTACTGAACTGCTTACTGGTGAAAAGGTCACTTCACCCGGCGAGGAATTTGATAAAGTGTTTACTGCCATCTGCGA GGGCAAGATTATCGACCCTATGATGGAGTGCTTGAATGAATGGAATGG- TGCCCCTATCCCGATTTGCTAG<u>TCTAAGAGACC</u>CC
AtC4H _G2	<u>GG</u><u>GGTCTCTACAA</u>ATGGATTTATTGTTACTCGAAAAGAGTTTGATTGCTGTGTTT- GTCGCCGTCATCTTGGCTACCGTTATCAGTAAATTGCGAGGAAAGAAGCTGAAATTGCCACCCGGACCAATTCCGATCC CCATTTTCGGCAATTGGCTGCAGGTGGGTGACGACCTGAATCATCGGAACTTGGTT- GACTATGCGAAGAAGTTTGGTGACTTATTCTTGCTGCGAATGGGACAAAGAAATTTGGTGGTGGTGTCTTCTCCTGACC TCACTAAAGAGGTCTTGCTGACCCAGGGAGTCGAATTCGGTAGCCGAACTCGGAATGTT- GTTTTCGATATCTTTACTGGTAAGGGCCAGGACATGGTTTTCACCGTCTATGGAGAACACTGGCGAAAGATGCGGCGAA TTATGACAGTCCCGTTCTTCACAAATAAGGTGGTGCAGCAAAACAGA- GAGGGATGGGAATTCGAGGCCGCCTCTGTCGTCGAGGACGTGAAGAAGAACCCTGACTCCGCGACCAAGGGCATAGTCC TTAGAAAGCGATTACAGCTTATGATGTACAATAACATGTTTCGAATTATGTTT- GACCGACGATTCGAATCTGAAGACGACCCCCTGTTCTTACGATTAAAAGCCCTCAACGGAGAACGTTCTCGCTTGGCCC AAAGTTTCGAATACAATTACGGCGACTTTATCCCCATTTTGCGGCCATTCTTACGAG- GATACCTTAAAATCTGCCAGGACGTTAAGGACCGGCGAATTGCCTTGTTTAAGAAATATTTCGTGGACGAACGGAAACA GATCGCTAGTTCCAAACCCACCGGATCAGAGGGCCTGAAGTGCGCAATCGACCA- TATTCTGGAAGCAGAACAGAAAGGTGAGATAAATGAAGATAACGTGCTCTATATTGTTGAAAATATAAACGTGGCTGCT ATCGAAACTACCCTGTGGTCGATAGAATGGGGTATCGCTGAACTCGTTAATCACCCAGA- GATTCAATCGAAACTGCGCAATGAGTTAGATACCGTGCTGGGCCCCGGAGTCCAGGTGACTGAACCCGACTTGCATAAG TTGCCGTATTTGCAGGCGGTTGTGAAAGAAACCTTGAGATTACGAATGGCCATCCCCTTGTT- GGTTCCGCATATGAATTTGCACGACGCTAAATTGGCCGGTTATGACATTCCTGCCGAGTCTAAGATTTTGGTCAACGCC TGGTGGCTCGCTAATAATCCAAATTCCTGGAAGAAGCCCGAG- GAATTCCGCCCCGAACGATTCTTCGAGGAAGAGAGCCATGTCGAGGCAAATGGCAACGATTTTCGATACGTTCCTTTCG GAGTGGGCCGAAGATCATGCCCTGGTATCATTCTGGCTCTGCCGATCCTCGGTAT- TACTATCGGACGAATGGTGCAAAATTTTGAATTGCTGCCGCCCCCGGGCCAATCGAAGGTCGACACCTCAGAAAAGGGC GGCCAGTTTTCCCTGCATATTTTGAATCATTCGATTATTGTCATGAAGCCTCGTAATTGCTAG<u>GGATAGAGACC</u>
AtCPR1 _G3	<u>GGG</u><u>GTCTCTCCA</u>CATGGCTACCATCCCTATGGACATTGTGAATGATATCTTTCTCAGATT- GCCCGCCAAGACCTTGGTCCGATGTCGGGCGTTGTCGAAACCTTGCTACCACTTGATTAATGACCCCGACTTTATTGAG TCGCACTTACACCGAGTGTTACAGACAGGCGACCACCTGATGATCCTGTT- GCGTGGCGCATTGAGATTGTATTCCGTGGACTTGGACTCTTTGGACTCGGTTTCGGATGTGGAACACCCAATGAAGAGA GGCGGACCAACAGAGGTCTTCGGATCCAGCAATGGCTTGATTGGTTT- GTCTAACTCCCCTACCGACCTTGCTGTGTTCAACCCCTCTACCCGCCAAATCCACCGGTTGCCACCCAGCTCAATTGAC CTGCCAGACGGAAGTTCTACAAGAGGCTATGTCTTCTATGGTTT- GGGCTACGACTCAGTGTCTGACGACTACAAGGTCGTTCGAATGGTGCAATTCAAAATCGACTCCGAGGACGAGCTCGGA TGCTCCTTCCCCTACGAAGTCAAGGTCTTCTCTCTGAAGAAGAACTCGTGGAAACGAATAGA- GAGCGTCGCAAGCTCTATCCAACTGCTGTTCTATTTCTACTACCACTTGTTATACAGAAGAGGCTACGGAGTGCTGGCC GGCAACAGCTTGCACTGGGTGTTGCCCCGACGACCGGGCCTTATCGCATTCAACCT- GATCGTGCGATTCGACTTGGCTCTCGAGGAATTCGAAATCGTCCGATTCCCAGAGGCTGTGGCCAACGGCAACGTGGAC ATCCAAATGGACATCGGCGTGTTGGACGGCTGCTTGTGTCTGATGTGCAAT- TACGACCAATCCTACGTGGACGTCTGGATGATGAAGGAGTATAACGTCCGAGACTCCTGGACAAAAGTGTTCACTGTGC AAAAGCCTAAGTCCGTCAAATCTTTCTCTTACATGCGACCCCTGGTTCACAGCAAAGA- CAAGAAGAAAGTCTTGCTGGAATTGAACAACACCAAATTGGTGTGGTTCGACTTGGAATCCAAGAAGATGTCAACCTTG CGCATCAAAGACTGTCCCTCTTCCTACTCCGCCGAGTTGGTCGTCTCGTCCTTGGTTT- TAGGCTGCAAGGGCGACTTGAACAATATCAAATACCGCAAGGAGCAGCAGGCCAAAGAGGCCCGGGAAGCGAAGATCAT GCAAAATACAAAGCGTAGAGGAGACTTCTTGTCAAAGGGTTTTAAATTGGTGTTGTAG<u>GTATAGAGACC</u>GG
At4CL2 _G2	<u>GG</u><u>GGTCTCTACAA</u>ATGACTACTCAGGACGTCATCGT- GAACGATCAAAACGACCAAAAGCATGCTCCAACGATGTTATCTTTCGTTCCCGTCTTCCCGACATTTATATTCCCAAT CATTTGCCCTTGCATGATTATATTTTCGAGAACATTTCTGAATTT- GCTGCAAAACCCTGTTTAATTAATGGCCCGACAGGAGAGGTGTATACATATGCTGACGTGCATGTCACCAGTCGTAAGT TGGCTGCGGGATTACACAATTTAGGAGTCAAACAGCATGATGTGGTCATGATACTGTT- GCCAAATAGCCCCGAGGTCGTTTTGACCTTCTTGGCAGCTTCTTTTATAGGTGCCATTACTACTTCTGCTAATCCATTC TTTTACCCCCGCCGAAATCTCCAAGCAGGCTAAGGCGAGCGCTGCAAAGTTAATTGTGACACA- GAGCCGATATGTTGACAAGATTAAGAATTTGCAGAATGATGGTGTGCTGATTGTTACTACTGATTCAGATGCAATTCCG GAGAATTGTTTGCGCTTTAGCGAATTGACACAAAGCGAGGAGCCTA- GAGTCGATTCTATTCCAGAAAAGATCTCTCCCGAGGATGTTGTTGCCTTACCCTTTTCGTCTGGAACCACCGGATTGCC TAAGGGCGTTATGTTGACCCATAAGGGATTAGTTACTTCCGTCGCTCAACAGGTGGATGGA- GAAAACCCTAACTTGTATTTTAATCGAGACGATGTCATCTTGTGCGTGTTACCCATGTTTCACATTTATGCATTGAATA GTATTATGTTGTGCTCTTTGCGCGTGGGAGCTACCATACTCATTATGCCCAAATTTGAGAT- TACCCTTCTCTTGGAACAAATTCAGCGATGCAAGGTGACCGTCGCCATGGTGGTTCCCCCTATTGTGTTGGCCATTGCC AAAAGCCCCGAAACCGAAAGTACGACTTATCTAGCGTCCGAATGG- TGAAATCAGGTGCGGCGCCCCTGGGCAAAGAATTAGAGGACGCAATATCTGCCAAATTCCCAAATGCGAAATTGGGCCA AGGTTACGGAATGACCGAGGCGGGCCCAGTCTTGGCTATGAGTTTGGGTTTCGCGAAG- GAACCCTTCCCCGTTAAATCCGGCGCCTGCGGCACCGTCGTCCGTAATGCTGAAATGAAAATTTTGGACCCTGATACCG GCGACTCCCTGCCGCGGAATAAGCCTGGTGAGATCTGTATTAGAGGAAATCAGAT- TATGAAGGGATACTTGAACGATCCTCTGGCTACCGCCAGCACCATTGACAAGGACGGATGGTTACATACAGGTGATGTG GGTTTCATTGACGACGATGATGAATTATTTATCGTCGACCGACTGAAGGAGTTGATTAAA- TATAAGGGCTTCCAGGTCGCGCCCGCCGAATTGGAAAGTCTTCTGATCGGCCACCCCGAGATTAACGACGTCGCCGTGG TGGCTATGAAAGAAGAGGACGCCGGAGAAGTCCCCGTGGCTTTCGTTGTCCGAAGCAAG- GACTCGAACATCTCTGAGGACGAGATTAAACAGTTTGTCAGCAAGCAAGTGGTTTTCTACAAACGCATTAATAAGGTCT

	TCTTTACCGATTCCATCCCAAAGGCCCCTAGTGGTAAAATTCTCCGAAAAGACTT- GCGAGCTCGCCTTGCTAACGGCTTGATGAATTAA<u>GGATAGAGACC</u>GG
At4CL3 _G1	<u>GGGGTCTCTAATG</u>ATTACAGCTGCCTTGCATGAGCCCCAAATCCATAAGCCCACTGACAC- CAGTGTGGTGTCGGACGACGTGCTCCCCCACTCCCCGCCCACCCCTCGGATCTTTAGAAGCAAATTGCCTGATATCGAT ATTCCCAATCATCTGCCCTTGCATACCTATTGTTTTGAGAAGCTGAGCAGCGTCTCT- GATAAACCATGCCTTATTGTGGGTAGCACTGGCAAGTCCTATACATATGGAGAGACTCATTTGATTTGCCGTCGCGTTG CGTCTGGTCTGTATAAACTCGGCATTCGAAAGGGAGATGTGATTATGATTTTGTTGCAGAA- TAGCGCAGAATTTGTCTTTTCGTTTATGGGAGCAAGCATGATTGGCGCTGTGTCTACAACTGCTAATCCCTTTTATACA TCGCAGGAATTGTACAAGCAATTGAAGAGCTCGGGTGCTAAGCTGATTATAACACA- TAGTCAGTATGTGGACAAGTTGAAGAATTTGGGCGAGAATTTGACTCTCATTACCACAGACGAGCCCACTCCAGAAAAC TGCTTGCCCTTTTCTACTCTTATTACTGATGATGAGACTAATCCCTTCCAGGAGACTGTGGA- CATTGGTGGCGATGACGCCGCTGCCCTGCCCTTTTCCTCGGGTACCACCGGTTTGCCCAAGGGTGTCGTCTTAACCCAT AAATCGTTGATTACTTCAGTGGCTCAGCAGGTGGACGGCGACAATCCTAACTTGTATTT- GAAGTCCAATGATGTTATTCTGTGTGTCTTGCCTCTGTTTCACATTTATTCCCTGAACTCGGTTTTATTGAACTCGTTG CGTTCGGGAGCAACTGTGTTGTTGATGCACAAATTCGAAATAGGAGCTTT- GCTCGACCTCATCCAGCGGCACCGCGTCACCATTGCTGCATTAGTCCCACCCTTAGTCATTGCCCTGGCCAAGAATCCC ACAGTCAACTCCTACGACCTTTCTTCAGTGCGCTTTGTGCTGTCTGGCGCGGCCCCCTT- GGGCAAGGAGTTACAGGACAGCTTGCGACGAAGATTGCCGCAGGCTATTTTGGGTCAAGGTTACGGCATGACTGAAGCG GGTCCTGTCCTCTCCATGTCTCTCGGCTTTGCAAAGGAGCCCATCCCTAC- TAAGTCCGGCAGCTGCGGCACCGTTGTTAGAAATGCTGAACTGAAGGTTGTGCATCTGGAAACTCGGTTGTCGTTGGGC TATAATCAGCCCGGCGAAATATGCATTAGAGGCCAGCAAATTATGAAGGAATATCTGAAT- GACCCTGAGGCAACAAGCGCGACCATCGATGAGGAAGGATGGTTGCATACTGGCGATATCGGTTACGTCGACGAGGACG ACGAAATCTTTATCGTCGACAGATTGAAGGAAGTTATTAAATTTAAGGG- TTTCCAAGTTCCCCCTGCGGAATTGGAATCCCTCTTGATTAACCATCACTCTATCGCTGACGCCGCCGTCGTGCCCCAG AACGACGAGGTGGCCGGCGAGGTCCCCGTTGCCTTTGTTGTCCGCAGTAACGGTAACGACAT- TACCGAAGAGGACGTGAAGGAGTACGTCGCGAAGCAAGTTGTCTTTTACAAACGACTGCATAAAGTGTTCTTCGTCGCC TCGATCCCAAAGTCCCCCTCAGGTAAGATCCTGCGAAAAGATTTGAAGGCCAAATTGTGCTAG<u>TCTAAGAGACC</u>CC
AtCHS3 _G2	<u>GGGGTCTCTACAAA</u>TGGTCATGGCAGGCGCCTCGAGTCTTGACGAAATACGA- CAAGCCCAACGAGCAGACGGTCCCGCTGGTATTTTAGCCATAGGTACCGCGAACCCAGAAAATCACGTTTTGCAGGCAG AATACCCCGATTATTATTTCCGAATAACAAATTCTGAGCATATGACTGATTTGAAAGA- GAAGTTTAAACGCATGTGTGATAAGTCCACCATCCGAAAGCGACATATGCACTTGACTGAAGAGTTCTTGAAAGAGAAT CCCCATATGTGCGCATATATGGCCCCCTCATTGGATACTCGACAAGATATT- GTCGTCGTGGAAGTGCCGAAACTTGGAAAGGAAGCCGCTGTCAAAGCGATTAAAGAATGGGGTCAACCTAAATCTAAAA TTACCCACGTGGTGTTTTGTACCACTTCAGGTGTTGATATGCCCGGCGCCGAT- TATCAACTGACAAAATTGTTGGGATTGCGACCGTCTGTTAAACGACTTATGATGTATCAACAGGGCTGTTTTGCAGGTG GCACAGTTTTGCGAATTGCCAAAGACTTGGCGGAAAATAATCGAGGTGCCCGAGTGCTT- GTCGTTTGTAGTGAAATTACTGCTGTCACATTTCGAGGACCTAGCGATACACATTTGGATAGCTTAGTTGGCCAAGCCT TGTTTAGCGACGGAGCTGCGGCTTTAATAGTCGGTTCTGATCCGGATACCAGTGTT- GGCGAAAAGCCTATATTCGAAATGGTTTCAGCTGCCCAAACTATTCTGCCCGATTCCGACGGAGCTATTGATGGCCATC TCCGAGAGGTCGGCTTGACATTTCACCTTTTAAAAGACGTCCCTGGATTGATTTCAAAGAA- TATCGTTAAATCTCTCGATGAGGCTTTCAAGCCACTGGGTATCTCTGATTGGAATTCTTTGTTTTGGATTGCTCATCCC GGCGGACCGGCTATTCTGGATCAAGTTGAAATTAAACTCGGTTTAAAAGAGGAAAAGAT- GAGAGCCACCCGGCATGTTTTATCTGAATACGGTAATATGTCTTCTGCCTGTGTCTTGTTTATTCTTGATGAAATGCGA CGTAAATCGGCCAAAGACGGCGTTGCGACAACTGGCGAGGGTTTAGAATGGGGAGTTT- TATTCGGCTTTGGCCCCGGCTTGACCGTCGAAACCGTGGTGCTTCATTCGGTGCCACTGTAG<u>GGATAGAGACC</u>
AtCHS1 _G1	<u>GGGGTCTCTAATG</u>TTAGTCTCAGCCCGAGTGGAAAAGCAGAAACGAGTGGCCTACCAAG- GAAAAGCCACCGTTCTGGCCTTGGGAAAAGCGCTGCCCTCGAACGTCGTGTCGCAAGAAAACTTGGTGGAAGAATACTT GCGGGAGATTAAGTGTGACAATTTGAGCATTAAGGATAAATTGCAGCATTTGTG- TAAGTCTACTACCGTGAAAACTAGATATACCGTGATGTCGAGAGAAACCTTACATAAGTATCCGGAGTTGGCTACTGAA GGCTCGCCCACTATTAAGCAACGGTTGGAAATCGCTAATGACGCCGTGG- TCCAAATGGCTACGAGGCCTCGCTCGTCTGTATTAAAGAGTGGGGCCGAGCGGTCGAGGACATTACCCACCTGGTGTA TGTCAGTAGTTCTGAATTTCGCCTGCCTGGTGGCGACCTCTATTTGTCCGCCCAATTGGGTT- TATCTAATGAAGTCCAACGGGTCATGTTATACTTCTTGGGTTGTTACGGCGGACTCTCTGGATTGCGAGTCGCTAAGGA TATCGCGGAAAATAATCCGGGCTCTCGGGTCCTCTTGACTACTAGCGAAACAACTGTGTT- GGGCTTTCGGCCCCCGAATAAGGCCCGACCATATAATCTTGTGGGCGCGGCTCTGTTCGGCGACGGCGCCGCCGCTTTG ATTATTGGCGCCGATCCCACTGAATCTGAGTCCCCCTT- TATGGAATTGCATTGCGCAATGCAACAATTTCTGCCGCAGACTCAAGGCGTCATTGATGGTCGTCTCAGTGAGGAAGGT ATCACTTTTAAATTGGGCCGAGATTTGCCGCAGAAAATTGAGGATAATGTTGAAGAATTTT- GTAAGAAATTGGTCGCCAAAGCCGGTTCGGGAGCTCTGGAATTAAACGATTTGTTTTGGGCCGTGCACCCCGGCGGCCC TGCGATTTTGTCCGGATTAGAAACTAAATTGAAATTGAAACCGGAGAAATTA- GAGTGTTCTCGTCGAGCCCTGATGGACTACGGTAATGTGTCTTCGAATACTATCTTTTATATTATGGATAAGGTGCGAG ACGAATTGGAAAAGAAGGGAACCGAAGGTGAGGAATGGGGATTGGGTTTGGCATTT- GGCCCCGGCATTACCTTTGAGGGTTTCTTGATGCGAAATCTGTAG<u>TCTAAGAGACC</u>CC
AtCHI1 _G3	<u>GGGGTCTCTCCACA</u>TGGGCACTGAAATGGTGATGG- TGCACGAAGTTCCCTTCCCGCCTCAAATTATTACCAGTAAACCCTTGTCCTTGTTGGGACAGGGAATTACCGATATAGA AATTCATTTCTTGCAGGTCAAATTTACCGCAATTGGCGTGTATTTGGACCCCTCT- GACGTCAAGACCCACTTGGACAATTGGAAGGGTAAGACTGGCAAGGAGTTGGCAGGTGACGACGATTTCTTTGATGCAT TGGCATCTGCTGAAATGGAGAAAGTCATCCGCGTCGTTGTCATCAAAGAAATTAAGGG- TGCCCAATATGGCGTCCAACTGGAAAACACAGTCCGCGACCGACTCGCAGAAGAAGACAAATATGAAGAGGAAGAGGAG ACAGAATTGGAGAAAGTGGTGGGATTCTTTCAATCGAAATATTTTAAGGCCAATTCAGTGAT- TACCTATCACTTTAGTGCAAAGGACGGAATCTGTGAAATTGGATTCGAAACCGAGGGAAAGGAAGAAGAAAAGTTGAAA

	GTTGAAAACGCTAACGTTGTTGGCATGATGCAACGGTGGTACCT- GAGTGGCTCTCGAGGCGTGTCCCCATCTACCATAGTGTCGATTGCCGACTCTATTTCGGCTGTGTTGACTTAG<u>GTATAG</u> <u>AGACCGG</u>
RtTAL _G1	GGGGATCC<u>GGTCTCT</u>AATGGCACCACGTCCAACAAGTCAAAGCCAAGCTCGCACTT- GTCCAACCACACAAGTTACACAAGTGGATATTGTGGAAAAGATGTTGGCTGCTCCCACTGATTCTACCTTGGAGTTGGA TGGATATTCCCTGAATCTGGGCGATGTGGTGTCTGCAGCCCGCAAAGGTCGTCCCGTGCGGG- TGAAAGATTCGGATGAAATTCGTTCGAAAATAGATAAGTCAGTTGAATTTCTGCGTTCCCAGTTGTCTATGTCCGTGTA TGGAGTGACTACCGGTTTCGGTGGTTCGGCTGATACTCGTACTGAAGATGC- TATTAGCCTTCAGAAAGCATTGCTGGAACATCAACTGTGTGGAGTGTTACCCAGCTCTTTTGATTCATTTCGATTAGGT CGAGGCTTGGAAAATAGCTTGCCTTTGGAAGTCGTGAGAGGAGCAATGACCATTCGTGT- GAATTCTCTGACTCGAGGACATTCTGCGGTGCGTTTGGTTGTGTTAGAAGCTTTAACTAATTTCTTAAATCATGGAATT ACACCGATTGTGCCACTGAGAGGAACTATTTCAGCCTCGGGTGATTTGAGTCCACTGAGCTA- TATCGCGGCTGCAATTTCCGGACATCCCGATTCTAAAGTCCATGTTGTGCATGAAGGAAAAGAGAAAATTTTGTATGCT CGAGAAGCTATGGCCTTGTTTAATTTGGAACCAGTTGTGTTAGGACCTAAAGAGGGCTT- GGGCCTGGTGAATGGAACTGCAGTGAGCGCTTCCATGGCTACTTTGGCCTTGCATGATGCCCATATGTTGTCCTTGTTG TCCCAATCTCTGACTGCTATGACCGTGGAAGCTATGGTTGGACATGCGGG- TTCTTTTCATCCTTTCTTACATGATGTGACTCGACCCCATCCAACCCAAATTGAGGTTGCTGGCAATATTCGAAAATTG TTGGAAGGCTCCCGGTTCGCAGTGCATCACGAAGAAGAAGTGAAAGTGAAAGATGATGAAG- GAATCTTGCGGCAAGATCGTTATCCGCTCCGGACCAGCCCCCAATGGTTGGGTCCCTTAGTGTCGGATTTGATACATGC GCATGCTGTGTTGACTATTGAAGCGGGACAATCCACAACAGATAATCCACTGATT- GATGTGGAAAATAAAACCTCTCATCATGGTGGTAACTTTCAAGCCGCGGCGGTCGCTAATACAATGGAAAAGACCAGAT TGGGCTTGGCGCAAATAGGTAAATTAAATTTTACTCAATTGACTGAAATGTTGAATGCGGG- TATGAATAGAGGTTTGCCGTCATGTCTGGCCGCTGAGGACCCTTCTCTGTCTTATCATTGTAAAGGTTTGGATATTGCA GCCGCTGCTTATACTTCTGAACTGGGTCATCTGGCAAATCCAGT- TACCCACCCACGTTCAACCTGCCGAAATGGCTAATCAAGCCGTGAATTCTTTGGCCTTGATTTCCGCCCGGCGAACCACA GAATCTAATGATGTGCTCTCGTTGCTGCTGGCTACACATCTGTATTGTGTGCTGCAGGCTA- TAGATCTGCGAGCTATTGAATTTGAATTTAAGAAACAATTTGGTCCTGCGATTGTGAGCTTGATTGATCAACATTTCGG ATCAGCGATGACTGGAAGCAATTTGCGTGATGAATTGGTGGAAAAGGTTAATAAAACCTT- GGCAAAACGATTGGAACAAACTAATTCCTATGATCTTGTTCCTAGATGGCATGATGCTTTCTCTTTTGCTGCGGGAACA GTGGTTGAAGTTTTGAGTAGTACATCTTTGTCTTTGGCTGCTGTGAATGCTTGGAAAGTT- GCAGCTGCGGAATCTGCAATATCATTAACACGACAGGTTCGAGAAACTTTCTGGAGTGCTGCCTCTACTAGTTCCCCAG CACTGTCATATTTAAGCCCACGAACACAAATTTTGTATGCTTTTGTGCGGGAAGAATTGGG- TGTTAAAGCGCGGCGTGGTGACGTGTTTCTGGGAAAACAGGAAGTTACCATTGGATCTAATGTTTCGAAAATTTACGAA GCTATTAAATCCGGACGGATTAATAACGTGTTGTTGAAAATGCTGGCCTAG<u>TCTATGAGACC</u>CCTAGGCC
FjTAL _G1	GGG<u>GTCTCT</u>AATGAACACCATCAACGAGTACCTGTCTCTGGAAGAGTTTGAGGC- TATTATCTTCGGAAACCAGAAGGTTACTATTTCCGATGTTGTCGTGAATCGAGTGAACGAGTCCTTCAACTTCCTTAAG GAGTTCTCAGGCAACAAGGTGATTTACGGCGTGAATACGGGATTT- GGCCCTATGGCCCAGTACCGAATCAAGGAGTCGGACCAGATTCAGCTGCAATACAACCTGATCCGATCCCACTCCTCGG GCACTGGCAAGCCCTTGTCCCCCGTTT- GTGCCAAAGCGGCAATCCTGGCTCGACTGAACACCCTTTCTCTGGGTAACAGCGGTGTGCATCCTTCTGTCATTAACCT GATGTCTGAGCTCATCAACAAGGACATCACCCCTCTGATTTTTGAACACGG- TGGCGTCGGAGCTTCCGGTGATCTGGTGCAGCTCAGTCATCTGGCATTGGTTCTCATCGGCGAGGGTGAGGTGTTTTAC AAGGGCGAACGGCGACCCACCCCTGAAGTCTTCGAGATT- GAGGGACTTAAGCCCATCCAGGTCGAGATCAGAGAAGGCCTCGCTCTCATCAATGGTACTTCCGTGATGACCGGTATCG GCGTCGTGAACGTCTACCACGCCAAGAAGCTGCTGGACTGGTCCCTGAAGTCGTCCTGTGC- TATTAACGAGCTGGTCCAGGCCTACGACGACCACTTTAGCGCCGAACTGAACCAGACCAAGCGGCATAAGGGCCAGCAG GAGATTGCCCTCAAGATGCGACAGAATCTTAGTGATAGTACACTGATCCGAAAGCGAGAG- GACCACCTCTACAGCGGCGAAAACACGGAAGAAATCTTCAAGGAAAAAGTGCAGGAGTACTATTCGCTCCGATGTGTGC CCCAGATTCTGGGACCCGTTCTGGAGACTATCAACAACGTGGCCTCCATCCTCGAGGAT- GAGTTCAACTCTGCTAACGACAATCCTATCATTGACGTGAAGAACCAGCACGTCTATCACGGAGGTAATTTCCACGGAG ACTACATCTCTCTCGAGATGGATAAGCTGACAGTTGATTACCAAGCTGACCATGCTT- GCCGAGCGACAACTCAACTACCTACTCAACTCTAAGATTAACGAGCTCCTTCCTCCGTTCGTGAACCTCGGCACTCTCG GCTTTAACTTCGGCATGCAGGGTGTTCAATTTACAGCTACCAGTACAACCGCCGAGTCCCA- GATGCTGTCCAACCCTATGTACGTGCACTCTATCCCCAACAATAATGACAACCAGGACATCGTGTCGATGGGCACCAAC TCCGCTGTGATCACTTCGAAGGTGATCGAGAACGCTTTCGAAGTCCTGGCTATCGAAATGAT- TACTATCGTGCAGGCCATTGACTACCTCGGCCAGAAGGACAAGATTTCGTCTGTGTCCAAGAAGTGGTACGACGAGATC CGAAACATCATTCCTACTTTTAAAGAGGACCAGGTCATGTACCCCTTCGTTCA- GAAGGTGAAGGACCACCTGATCAACAACTAA<u>TCTAAGAGACC</u>CC
HaTAL _G1	GGG<u>GTCTCt</u>AATGCGACACCAGGTGACCTTGACCGGTGCAGGCCTCACTATCGAG- GACGTCGTCCGGGTCGCCCGGCACCATCAGCCTGTGGGTCTGACTGACAACCCCGAAATCCTTCAGCGAATTGAGGACT CATGTGCATATATCAATGACGCTGTGAAGGCCTCCAAGCCTGTCTACGGTGTTACCA- CAGGATTTGGTGGTATGGCCGATGTGGTGATTAGCAGCGAGGAGGCCGCCGACCTCCAAAACAATGCCATCTGGTATCA CAAGACCGGTGCCGGCAAGCTGCTGCCTCTAGCT- GATGTCCGAGCCGCCATGCTGCTGCGGGCCAACTCACACATGCGAGGTGTCTCTGGTATCCGACTCGAGATTATCCAGC GTATGATGACATTCCTGAACGCCAACGTCACCCCCCACGTACGA- GAATTCGGCTCCATCGGAGCTTCAGGTGATCTGGTGCCGCTCATTAGCATTACAGGAGCCCTGCTAGGTACCGATCCTG CCTTTAGAGTCGACTTCGACGGAGAGAACATCGACTGCCTGGAGGCTCTCGAACGGCT- GAACCTGCCACGACTGGAGCTCCTACCCAAGGAAGGACTGGCTATGATGAACGGTACCTCGGTGATGACGGGCATCGCC TCCAACGTGCTGCACGACGCACGAATTCTGTTGGGTTTGGCTCTGAACATCCACGGCCTGAT- GATTCAAGGTCTGCAGGGTACTAACCAGTCATTCCACCCCTTCATTCACCAGCACAAGGCTCACACCGGCCAAGTGTGG GCTGCCGATCACATGCTGCAGATCCTCGAGGGTTCAGCACTCTCTCGAGAT- GAATTGGACGGCCGCCATGAGTACCGAGAGGGAGACTTGATTCAGGACCGGTACAGTTTGAGATGCCTGCCTCAATTCC

```
                    TGGGACCTATTATTGACGGTATGGCCTACATCACCCACCATCTTCGGGTGGA-
                    GATCAACTCGGCAAACGACAATCCCCTAATTAACACCGAGGCTGGAGCATCCTACCATGGTGGTAATTTTCTGGGACAG
                    TACATCGGCGTGGGCATGGACCAACTCCGTTATTACATGGGCCT-
                    GATGGCCAAGCACCTGGACGTGCAGATTGCTCTGCTGGTTTCGCCCCAGTTCAACAATGGCCTCTCGGCTTCGCTGGTG
                    GGCAACACCGACCGGAAGGTGAACATGGGACTCAAGGGCCTTCAGAT-
                    TAGCGGCAACTCCATCATGCCTATTTTGGGATTCCTTGGAAATTCTCTCGCTGACAGATTCCCCACCCACGCCGAGCAG
                    TTTAACCAAAACATCAACAGCCAGGGATTCGGAAGTGCCAACCTGGCCCGACAGACCATCGA-
                    GACTCTCCAACAATACATCGCCATTGCCCTCATTTTCGGCGTGCAAGCTGTAGATCTGCGCACCTTCAAGCGGACTGGT
                    CATTACAACGCCGTCGAGACACTGTCTCCCATGACAGCTAAGCTCTACTCCGCCATGCGA-
                    GAAGTCGTGGGAAAACCTATCTCTCACGAGCGACCTTACATTTGGAACGATAACGAGCAGGCTCTGGAGCAGCATATCT
                    CTGCTATCGTTAGCGACATTACCAACGACGGAATCATCCCCCAGGC-
                    TATCCAGGAAACCCTGGATTCTCTGCGATCTATTATCCTCTTCGCTTAATCTAaGAGACCCC
```

PhCHS _G3

```
                    GGGGATCCGGTCTCTCCACATGGTGACTGTCGAGGAGTATAGAAAGGCGCAAAGAGCTGAAG-
                    GACCCGCTACTGTCATGGCTATTGGAACAGCAACTCCGACAAACTGTGTGGATCAAAGTACCTACCCCGACTATTATTT
                    CAGAATCACTAACAGCGAGCACAAGACTGATTTAAAGGAGAAGTTCAAGAGAATGTGT-
                    GAGAAATCTATGATTAAGAAACGGTACATGCACCTGACAGAGGAAATCTTGAAAGAGAATCCATCTATGTGCGAATACA
                    TGGCTCCCTCTCTCGATGCTCGGCAAGACATTGTGGTGGTCGAGGTTCCCAAAT-
                    TAGGCAAAGAGGCTGCACAAAAGGCGATCAAGGAATGGGGTCAGCCAAAGTCTAAAATTACCCATTTGGTGTTTTGCAC
                    CACCAGCGGCGTTGACATGCCGGGATGT-
                    GACTATCAACTGACAAAGCTTCTCGGATTAAGACCATCGGTCAAGCGTCTTATGATGTACCAACAGGGATGCTTTGCCG
                    GTGGTACCGTTCTTCGGTTAGCTAAGGACCTTGCTGAAAACAACAAGGGTGCTCGAGTTCTT-
                    GTTGTTTGTAGTGAAATCACCGCTGTTACATTCCGAGGCCCCAATGACACCCATTTGGATAGTTTGGTTGGTCAAGCAT
                    TGTTTGGTGATGGTGCCGGCGCAATCATTATCGGCTCCGATCCTATTCCCGG-
                    TGTCGAACGTCCGCTTTTCGAGCTCGTTTCGGCAGCCCAAACTCTTCTCCCCGATAGCCATGGAGCGATTGATGGCCAT
                    TTGCGGGAAGTCGGACTTACTTTCCACTTGCT-
                    GAAAGATGTTCCTGGACTGATCAGTAAGAACATTGAAAAGAGTTTAGAGGAAGCTTTTCGTCCGTTGTCTATTTCTGAT
                    TGGAACTCTCTCTTCTGGATTGCTCATCCTGGCGGCCCTGCAATTCTCGACCAAGTGGAAAT-
                    TAAGTTGGGTTTGAAGCCCGAGAAACTCAAGGCTACCAGAAATGTGTTATCAAACTATGGCAACATGTCTTCGGCTTGT
                    GTCTTGTTTATTCTGGATGAAATGCGAAAGGCCTCTGCCAAGGAAGGTCTGGGCACTA-
                    CAGGCGAAGGATTAGAGTGGGGCGTGCTTTTCGGATTTGGTCCTGGCCTTACTGTGGAGACAGTTGTCCTTCACTCAGT
                    TGCTACTTAAGTATTGAGACCCCTAGGCC
```

MsCHI _G2

```
                    GGGGATCCGGTCTCTACAATGGCTGCCAGCATTACCGCAATAACCGTTGAAAATTTGGAG-
                    TATCCCGCAGTCGTGACTAGCCCTGTTACAGGAAAGAGCTACTTCTTGGGTGGAGCCGGCGAACGTGGATTAACTATTG
                    AGGGTAATTTTATTAAATTTACCGCAATTGGTGTGTACCTCGAGGACATAGCCGTTGC-
                    TAGTCTTGCAGCCAAGTGGAAAGGAAAGTCCAGCGAAGAATTGCTCGAAACTCTGGATTTCTATAGAGATATTATATCG
                    GGTCCATTCGAGAAACTCATCCGCGGAAGTAAAATCCGAGAGCTCTCGGGGCCCTGAA-
                    TATTCGCGGAAAGTGATGGAAAATTGCGTGGCCCATCTCAAGTCAGTCGGCACCTACGGCGACGCCGAGGCCGAGGCTA
                    TGCAGAAATTCGCCGAGGCTTTTAAACCGGTCAACTTCCCACCCGGCGCCTCAGTCTTT-
                    TATCGTCAGTCTCCTGACGGTATCCTCGGCTTGTCCTTTTCGCCGGACACTTCCATCCCTGAGAAAGAGGCAGCTTTGA
                    TAGAAAATAAAGCTGTTTCTTCGGCTGTTCTGGAAACAATGATT-
                    GGCGAACATGCCGTGTCCCCCGACTTGAAACGGTGCTTGGCGGCTCGGCTGCCTGCTTTGTTAAATGAAGGTGCCTTTA
                    AAATCGGAAATTAGGGGATTGAGACCCCTAGGCC
```

CrCPR _G2

```
                    GGGGATCCGGTCTCTACAATGGACAGCTCATCCGAAAAGTTATCTCCCTTT-
                    GAGCTCATGTCTGCTATTCTGAAAGGCGCAAAGCTGGACGGAAGTAATAGCTCCGACAGCGGTGTGGCCGTTTCCCCCG
                    CTGTCATGGCCATGCTCCTCGAAAACAAAGAATTGG-
                    TTATGATCTTAACCACCTCTGTCGCCGTCCTGATTGGATGCGTGGTGGTCCTTATTTGGCGCAGAAGTTCGGGCAGCGG
                    AAAGAAGGTTGTCGAACCCCCAAAATTGATCGTCCCCAAGTCCGTGGTCGAGCCCGAAGA-
                    GATCGACGAGGGCAAGAAGAAGTTCACTATTTTCTTCGGCACCCAGACCGGTACCGCCGAGGGATTTGCGAAAGCCTTG
                    GCAGAAGAGGCTAAGGCCCGCTACGAGAAAGCCGTGATTAAGGTCATCGA-
                    CATTGACGACTACGCCGCCGACGACGAGGATGTGAAGAAAAGTTTCGAAAGGAAACACTGGCCTTCTTTATTCTTGCG
                    ACCTACGGCGACGGAGAACCCACTGATAACGCAGCTCGGTTTTATAAGTGG-
                    TTCGTCGAAGGCAACGACCGAGGAGATTGGTTGAAGAACTTGCAGTACGGCGTGTTCGGATTGGGCAATCGACAGTACG
                    AACACTTTAATAAAATCGCAAAGGTCGTCGACGAAAAGGTCGCCGAGCAAGGTGGAAAACG-
                    TATAGTGCCTCTGGTCTTGGGTGATGACGATCAATGTATCGAGGACGATTTCGCAGCCTGGAGAGAAAACGTTTGGCCC
                    GAACTCGACAATCTCCTGAGAGATGAAGACGACACCACCGTCAGCACTACATATA-
                    CAGCCGCGATCCCTGAGTACCGCGTCGTCTTTCCCGATAAGTCGGACTCTTTAATCTCTGAGGCCAACGGTCACGCTAA
                    CGGCTACGCCAACGGTAATACAGTCTACGACGCTCAACACCCCTGTCGCTCCAACGTGGCGG-
                    TTAGAAAAGAATTGCACACCCCCGCCTCGGACAGATCCGTGTACACACTTAGACTTCGATATCGCCGGAACCGGTTTGTC
                    GTACGGCACCGGCGACCACGTCGGTGTCTATTGCGACAACTTGTCCGAGA-
                    CAGTCGAAGAAGCGGAACGACTCCTCAACTTGCCACCTGAGACATACTTTAGTTTGCACGCCGACAAGGAAGACGGCAC
                    TCCCTTGGCCGGAAGTTCCCTCCCTCCCCGTTTCCGCCATGCACCTTGCGCACTGCTCT-
                    GACACGATACGCCGACTTGTTGAACACCCCCAAGAAATCCGCCCTGTTGGCCTTGGCCGCATACGCCAGTGACCCCAAC
                    GAAGCAGACCGATTGAAGTACTTGGCCTCACCTGCTGGCAAAGACGAG-
                    TACGCCCAATCTCTCGTCGCTAACCAACGATCGCTTTTGGAAGTGATGGCCGAGTTCCCCTCGGCTAAACCGCCCTTGG
                    GTGTGTTCTTCGCTGCCATCGCACCCCGGTTGCAGCCGCGGTTT-
                    TACTCGATTTCCAGCTCACCCAGAATGGCGCCCTCGCGTATCCACGTTACCTGCGCGTTGGTGTACGAGAAGACCCCAG
                    GCGGTCGCATCCATAAAGGCGTCTGCTCCACCTGGAT-
                    GAAGAACGCAATCCCCCTGGAAGAGTCTCGCGATTGTTCTTGGGCCCCGATTTTCGTGCGTCAAAGTAATTTTAAGTTG
                    CCCGCTGACCCCAAGGTTCCAGTCATTATGATTGGTCCCGGCACAGGTTTGGCACCAT-
                    TTCGGGGCTTTCTCCAAGAGCGGTTGGCCTTGAAAGAAGAGGGTGCCGAGTTGGGAACAGCCGTCTTCTTCTTCGGCTG
                    TCGAAATCGAAAGATGGACTATATTTACGAGGACGAACTCAATCACTTCTTGGA-
                    GATCGGCGCCTTGTCAGAATTGCTGGTCGCCTTTTCCCGAGAAGGCCCTACAAAACAATACGTCCAGCATAAAATGGCT
```

	GAGAAAGCCAGCGACATCTGGCGCATGATCTCGGACGGCGCTTACGTGTATGTTTGTGG- TGACGCTAAGGGAATGGCACGAGATGTGCATCGAACATTGCATACTATCGCACAGGAACAAGGCTCTATGGACTCTACC CAAGCAGAAGGATTCGTTAAGAACTTGCAGATGACAGGTAGATACCTGAGAGATGTGTGG- TAG<u>GGATTGAGACC</u>CCTAGGCC
VvSTS1 _G3	<u>GGTCTCTCCAC</u>AATGGCCTCTGTGGAGGAGTTCAGAAACGCTCAGCGAGCCAAGGG- TCCCGCCACTATCCTTGCAATCGGCACTGCCACTCCCGACCACTGCGTGTACCAGTCTGACTACGCTGATTACTACTTC AAGGTTACAAAGTCCGAGCACATGACCGCCTTGAAGAAGAAGTTCAACCGTATCTGCGA- CAAGTCTATGATCAAGAAGCGGTACATTCACCTTACCGAGGAGATGCTCGAGGAGCATCCGAACATCGGAGCCTACATG GCCCCCTCGCTCAACATCCGACAGGAGATCATCACTGCCGAGGTCCCCAAGCTGGGTAAG- GAGGCCGCACTGAAGGCTTTGAAGGAGTGGGGTCAGCCCAAGTCCAAGATTACCCATCTGGTGTTTTGCACAACCTCTG GAGTGGAGATGCCTGGCGCCGACTA- CAAGCTCGCAAACCTCCTGGGCCTGGAGCCCAGCGTGCGACGGGTGATGCTGTACCACCAGGGATGTTACGCAGGCGGC ACTGTTCTGCGAACCGCCAAGGACTTGGCCGAGAACAACGCCGGAGCCCGAGTCCTT- GTCGTGTGCTCTGAAATTACCGTTGTCACTTTCCGAGGTCCCTCTGAGGACGCCCTCGATTCTCTCGTTGGTCAGGCAC TATTTGGTGATGGATCTGCCGCTGTCATTGTCGGCAGTGACCCT- GATATCTCGATCGAGCGACCCCTGTTTCAGCTTGTGTCAGCAGCACAGACCTTTATTCCCAACAGTGCCGGCGCCATCG CCGGCAACCTGAGAGAAGTTGGGCTGACCTTTCACCTGTGGCCTAACGTCCCCACCCT- GATTTCCGAAAACGTGGAGAAGTGTCTTACACAGGCCTTCGACCCGCTGGGCATTAGCGACTGGAACTCACTGTTCTGG ATCGCCCATCCCGGTGGCCCTGCCATCCTCGATGCCGTGGAGGCCAAGCTTAACCTCGA- CAAGAAGAAGCTCGAGGCCACCCGACACGTGCTGTCGGAGTACGGCAACATGTCCTCTGCTTGCGTGCTGTTTATCCTG GACGAGATGCGAAAAAAGTCCCTGAAGGGAGAGCGAGCTAC- CACCGGCGAGGGACTGGACTGGGGAGTGCTGTTCGGCTTCGGACCTGGCCTGACAATTGAGACTGTCGTCCTGCACAGC ATTCCCATGGTGACCAACTAA<u>GTATAGAGACC</u>
RtSTS _G3	<u>GGTCTCTCCAC</u>AATGGCCCCTGAAGAGTCTAGACACGCTGAGACTGCCGT- GAACCGAGCCGCTACAGTGCTGGCCATTGGCACGGCCAACCCCCCCAACTGCTACTACCAGGCCGACTTCCCCGACTTC TACTTTAGAGCTACCAACTCTGACCACCTGACCCACCTCAAGCAGAAGTTCAAGA- GAATCTGCGAGAAGTCCATGATCGAGAAGCGATACCTGCATCTTACTGAAGAGATTCTGAAGGAGAACCCTAACATTGC TTCTTTCGAGGCTCCCTCCCTCGATGTGCGACACAACATTCAGGTGAAGGAGGTGG- TGCTTCTGGGAAAGGAGGCGGCTCTGAAGGCGATTAACGAATGGGGACAGCCCAAGTCGAAGATTACCCGACTGATCGT TTGCTGCATCGCCGGCGTGGACATGCCTGGTGCAGACTACCAGCT- TACCAAGCTCCTGGGCCTGCAGCTCTCCGTCAAGCGATTTATGTTCTACCACCTGGGATGTTACGCCGGCGGAACCGTG CTGCGGTTGGCCAAGGATATTGCCGAGAACAACAAGGAGGCTCGGGTTCTGATTGT- GAGGTCCGAGATGACTCCGATCTGCTTCCGAGGCCCCAGCGAGACTCATATCGACTCCATGGTGGGACAGGCTATCTTC GGAGATGGTGCCGCAGCCGTGATCGTGGGCGCTAACCCCGATTT- GTCCATTGAACGACCAATCTTCGAGCTCATCTCAACATCTCAGACCATCATTCCCGAATCCGACGGCGCTATCGAGGGC CACCTGCTCGAGGTGGGCCTGTCATTCCAGCTGTACCAGACAGTTCCCTCTCTCAT- TTCCAACTGTATCGAAACTTGTCTGTCTAAGGCCTTTACTCCTCTAAACATTTCTGACTGGAACAGCCTCTTCTGGATT GCCCACCCCGGCGGCCGAGCAATCCTGGACGACATCGAGGCCACCGTCGGACTGAAGAAGGA- GAAGCTTAAGGCCACCCGCCAGGTGCTCAACGACTACGGAAACATGTCGAGCGCCTGTGTCTTCTTCATTATGGACGAG ATGCGAAAGAAGTCTCTCGCTAACGGACAGGTCACCACCGGCGAGGGCCT- GAAGTGGGGAGTGCTTTTCGGATTTGGACCCGGAGTGACTGTGGAGACAGTCGTTCTCTCTTCCGTGCCCCTGATCACT TAA<u>GTATAGAGACC</u>

6.1.3   Table of strains

In E. coli, GGV indicates the <u>G</u>olden <u>G</u>ate donor <u>V</u>ector and GGA the <u>G</u>olden <u>G</u>ate <u>A</u>ssemblies that are ready to be transformed in YL as expression cassettes. All the expression cassettes, except otherwise state, have zeta insertion sequences, pTEF as promoter and tLip2 as terminator.

E. coli strains		
Name	Genotype	Antibiotic Resistance
	Cas9	
JME4307	TOPO-PUraTgsy. PT cassette for GSY deletion.	KanR
JME4640	GGA_InsUp/DownGSYcas9CutSite100bp-NA-Tex-pTEF-RedStar-Tlip-	AmpR
JME4641	GGA_InsUp/DownGSYcas9CutSite1000bp-NA-Tex-pTEF-RedStar-Tlip	AmpR
JME4645	GGA_InsUp/DownGSYcas9CutSite100bp-NA-Tex-pTEF-XylosePathway-Tlip	AmpR
JME4646	GGA_InsUp/DownGSYcas9CutSite1000bp-NA-Tex-pTEF-XylosePathway-Tlip	AmpR
JME4647	GGA_IsUp/DownGSYcas9CutSite100bp(SfiI) -NATex-pTEF-Carotenoids pathway-Tlip	AmpR
JME4648	GGA_InsUpGSYcas9CutSite1000bp-NATex-pTEF-Carotenoids pathway-Tlip	AmpR
JME4652	ylReplicativeVector + GGA RedStar-InsGSYcas9CutSite100bp_clone12	AmpR
JME4653	ylReplicativeVector + GGA RedStar-InsGSYcas9CutSite1000bp_clone19	AmpR
JME4654	ylReplicativeVector + GGA CarotenoidsPathway-InsGSYcas9CutSite100bp_clone4	AmpR
JME4655	ylReplicativeVector + GGA CarotenoidsPathway-InsGSYcas9CutSite1000bp_clone4	AmpR
GGE0305	GGA_RS-URA-InsGSYCAS9CutSite1000	AmpR
GGE0306	GGA_RS-NoMarker-InsGSYCAS9CutSite1000	AmpR
GGE0307	GGA_RS-URA-InsGSYCAS9CutSite100	AmpR
GGE0308	GGA_RS-NoMarker-InsGSYCAS9CutSite100	AmpR
GGE0318	GGA_Xylose-NoMarker-InsGSYCAS9CutSite1000	AmpR
	Cpf1	
JME4990	Cpf1-Leu-RFPplatform	AmpR
JME4989	Cpf1-Hph-RFPplatform	AmpR
JME5232	Cpf1-Leu-gGSY	AmpR
JME5236	Cpf1-Hph-gGSY	AmpR
	AAA pathway	
GGE0178	GGV TOPO ARO1_G1	KanR

GGE0179	GGV TOPO ARO2_G2	KanR
GGE0316	GGV TOPO Aro4_G1	KanR
GGE0212	GGV TOPO ARO4 K221L_G1	KanR
GGE0188	GGV TOPO scARO4K229L-OptYL_G1	KanR
GGE0314	GGV TOPO Aro7_G2	KanR
GGE0181	GGV TOPO ARO7G139S_G2	KanR
JME4770	GGV TOPO scARO7T226I_G2	KanR
JME4802	GGV TOPO scAro3K222L-OptYL_G2	KanR
GGE 0060	GGV TOPO ARO8_G1	KanR
GGE 0061	GGV TOPO ARO10_G2	KanR
JME4966	GGA_URA3ex-Aro4	AmpR
JME4967	GGA_URA3ex-Aro4	AmpR
JME4920	GGA_LEU2ex-Aro7	AmpR
JME4921	GGA_LEU2ex-Aro7G139S	AmpR
JME4922	GGA_LEU2ex-Aro4K221L	AmpR
JME5007	GGA_URA3ex-Aro4K221L	AmpR
JME5008	GGA_LEU2ex-Aro4	AmpR
JME5010	GGA_URA3ex-Aro7wt	AmpR
JME5011	GGA_URA3ex-Aro7G139S	AmpR
JME5009	GGA_URA3ex-scAro4K229L	AmpR
JME5012	GGA_LEU2ex-scAro7T226I	AmpR
JME5013	GGA_URA3ex-scAro7T226I	AmpR
JME4998	GGA-LEU2ex-scAro3K222LOptYL	AmpR
JME5001	GGA_Nat-scAro3K222LOptYL	AmpR
JME5003	GGA_Hph-scAro3K222LOptYL	AmpR
GGE0192	GGA_URA3ex_Aro1-Aro2	AmpR
JME4987	GGA_Hph-Aro1-Aro2	AmpR
GGE0325	GGA_URA3ex_Aro4wt_Aro7wt	AmpR
JME5156	GGA_Nat-Aro4wt-Aro7wt	AmpR
GGE0230	GGA_URA3ex_Aro4K221L-Aro7G139S	AmpR
JME4786	GGA_URA3ex_scAro4K229L-scAro7T226I	AmpR
JME4839	GGA_LEU2ex-scAro4K229L-scAro7T226I	AmpR
JME4862	GGA_LEU2ex-scAro4K229L-scAro3K222L	AmpR
JME4838	GGA_URA3ex-scAro4K229L-scAro3K222L	AmpR
JME4986	GGA_Hph-scAro4K229L-scAro3K222L	AmpR
JME4863	GGA_URA3ex-Aro8-Aro10	AmpR
JME4940	GGA_Nat-Aro8-Aro10	AmpR
JME3284	JMP62_URA3 -TKL1	KanR
JME3286	JMP62_Leu2ex -TKL1	KanR
GGE0153	GGV TOPO Promoter region Aro8	KanR
GGE0155	GGV TOPO Promoter region Aro9	KanR
GGE0154	GGV TOPO Promoter region Aro10	KanR
	PVA	
GGE0119	GGV TOPO VioA_ G1	KanR
GGE0120	GGV TOPO VioB_G2	KanR
GGE0123	GGV TOPO VioE_G3	KanR
GGE0134	GGA URA3ex_VioA-VioB-VioE	AmpR
	Naringenin	

JME4830	GGV TOPO atPAL(at2g37040)OptYL_G1	KanR
JME4803	GGV TOPO atCHS3(at5g13930)OptYL_G1	KanR
JME4804	GGV TOPO at4CL3(at1g65060)OptYL_G1	KanR
JME4805	GGV TOPO atCPR1(at4g12560)OptYL_G3	KanR
JME4806	GGV TOPO atCHI1(at5g05270)OptYL_G3	KanR
JME4807	GGV TOPO at4CL2(at3g21240)OptYL_G2	KanR
JME4808	GGV TOPO atCHS1(at4g00040)OptYL_G1	KanR
JME4809	GGV TOPO atC4H(at2g30490)OptYL_G2	KanR
GGE0385	GGV TwistVector RtTAL-OptYL_G1	KanR
GGE0386	GGV TwistVector MsCHI-OptYL_G2	KanR
GGE0387	GGV TwistVector PhCHS-OptYL_G3	KanR
GGE0388	GGV TwistVector CrCPR-OptYL_G2	KanR
JME5151	GGA_URA-AtPAL	AmpR
JME5153	GGA_URA-At4Cl3	AmpR
JME4981	GGA_NAT-atPAL-atC4H-atCPR	AmpR
JME4919	GGA_NAT- atCHS1-atCHS3-atCHI	AmpR
JME4983	GGA_hph-atCHSI-at4CL2-atCHI	AmpR
JME5014	GGA_hph_atCHSI-at4CL2-atCHI1	AmpR
JME5121	GGA URA-at4Cl3-atC4H	AmpR
JME5126	GGA_Hph-at4Cl3-atC4H	AmpR
JME4865	GGA_Ura-at4CL3-at4CL2	AmpR
JME4866	GGA_Leu-at4CL3-at4CL2	AmpR
JME5122	GGA_Hph-at4Cl2	AmpR
JME5124	GGA _Ura-at4Cl2	AmpR
JME5143	GGA_Leu-RtTAL-MsCHI-PhCHS	AmpR
JME5144	GGA_ Nat-At4CL3-MsCHI-PhCHS	AmpR
JME5155	GGA _Leu-At4Cl3-MsCHI-PhCHS	AmpR
JME5158	GGA Nat-PhCHS	AmpR
JME4912	357- Naringenin Biosensor Vector	AmpR
	Resveratrol	
GGE0217	GGV TOPO FjTAL-OptYL_G1	KanR
GGE0241	GGV TwistPlasmid HaTAL-OptYL_G1	KanR
GGE0242	GGV TwistPlasmid At4CL2-OptYL_G2	KanR
GGE0243	GGV TwistVector VvSTS1-OptYL_G3	KanR
GGE0244	GGV TwistVector G3-Rt-STS-OptYL_G3	KanR
GGE0299	GGV TOPO HaTAL-OptYL_G1	KanR
GGE0300	GGV TOPO At4CL2-OptYL _G2	KanR
GGE0301	GGV TOPO VvSTS1-OptYL _G3	KanR
GGE0383	GGV TOPO Rt-STS-OptYL_G3	KanR
GGE0319	GGA_URA3ex_HaTAL_At4CL2_VvSTS1	AmpR
GGE0322	GGA_LEU2ex_HaTAL_At4CL2_VvSTS1	AmpR
JME4985	GGA_Hph-haTAL-at4CL2-VvSTS1	AmpR
JME5104	GGA LEU2ex-HaTal-At4CL-VvSTS	AmpR
	Melanin	
JME5111	CrCas9-LEU2-g4HPPD	AmpR
JME5116	CrCas9-HPH-g4HPPD	AmpR
JME5117	GGA LEU2-4HPPD	AmpR

JME5119	GGA HPH-4HPPD		AmpR

	Y. lipolytica strains.		
Name	Genotype	Auxotrophy	Reference
JMY195, Po1d	*MATA ura3-302 leu2-270 xpr2-322*	U-L-	(Barth and Gaillardin 1996)
JMY 2101	JMY195 + *LEU2*	U-	Dulermo R. *et al.* 2017
	Chassis strain and Ehrlich metabolites biosynthesis		
JMY7891	Po1d + *Nat-ARO8-ARO10*	U-L-	This work
JMY7892	Po1d + *URA3-ARO1-ARO2* + Nat-*ARO8-ARO10*	L-	This work
JMY8074	Po1d + *URA3-ARO4-ARO7* + *Nat-ARO8-ARO10*	L-	This work
JMY8070	Po1d + *URA3-ARO4K221L-ARO7G139S* + Nat-ARO8-ARO10	L-	This work
JMY7902	Po1d + LEU-scARO4K229L-scARO7T226I + Nat-ARO8-ARO10	U-	This work
JMY8306	Po1d + URA3-ARO1-ARO2 + ARO4-ARO7 + Nat-ARO8-ARO10	L-	This work
JMY8000	Po1d + LEU2-ARO1-ARO2 + URA3-ARO4K221L-ARO7G139S + Nat-ARO8-ARO10	Prototroph	This work
JMY7903	Po1d +URA3-ARO1-ARO2 + LEU-scARO4K229L-scARO7T226I + Nat-ARO8-ARO10	Prototroph	This work
JMY8109	Po1d + URA3-ARO1-ARO2 + scARO4K229L-scARO7T226I + Nat-scARO3K222L + HPH-ARO8-ARO10	L-	This work
JMY8032	Po1d + URA3-ARO1-ARO2 + scARO4K229L-scARO7T226I + LEU2-TKL+ Nat-ARO8-ARO10	Prototroph	This work
JMY8054	Po1d URA3-ARO1-ARO2 + scARO4K229L-scARO7T226I + scARO3K222L + LEU2-TKL+ NAT-ARO8-ARO10	Prototroph	This work
JMY7737	Po1d + URA3-scARO4K229L-scARO3K222L	L-	This work
JMY7906	Po1d + URA3-scARO4K229L-scARO3K222L + Nat-ARO8ARO10	L	This work
JMY7900	Po1d + URA3-ARO1-ARO2 + LEU2-scARO4K229L-scARO3K222L	Prototroph	This work
JMY8002	Po1d + URA3-ARO1-ARO2 + LEU2-scARO4K229L-scARO3K222L + Nat-ARO8ARO10	Prototroph	This work
JMY7481	Po1d + URA3-ARO1-ARO2	L-	This work
JMY7521	Po1d + URA3-ARO4-ARO7	L-	This work
JMY8250	Po1d + URA3-ARO1-ARO2 + LEU2-ARO4-ARO7	Prototroph	This work
JMY7735	Po1d +URA3-ARO1-ARO2 + LEU-scARO4K229L-scARO7T226I	Prototroph	This work
JMY8175	Po1d + URA3-ARO1-ARO2 + scARO4K229L-scARO7T226I + NAT-scARO3K222L	L-	This work
	Strains for feedback regulation test.		
JMY8073	Po1d + URA3-ARO4 + NAT-ARO8ARO10	L-	This work
JMY8077	Po1d + URA3-ARO4K221L + NAT-ARO8ARO10	L-	This work
JMY8079	Po1d + URA3-ARO7 + NAT-ARO8ARO10	L-	This work
JMY8081	Po1d + URA3-ARO7G139S + NAT-ARO8ARO10	L-	This work

JMY8354	Po1d + URA3-ARO1-ARO2 + LEU-ARO4 + NAT-ARO8ARO10	Prototroph	This work
JMY8356	Po1d + URA3-ARO1-ARO2 + LEU2-ARO4K221L + NAT-ARO8ARO10	Prototroph	This work
JMY8358	Po1d + URA3-ARO1-ARO2 + LEU2-ARO7 + NAT-ARO8ARO10	Prototroph	This work
JMY8360	Po1d + URA3-ARO1-ARO2 + LEU2-ARO7G139S + NAT-ARO8ARO10	Prototroph	This work
	Naringenin		
JMY8030	Po1d + NAT-AtPAL-AtC4H-AtCPR + HPH-AtCHSI-At4CL2-AtCHI1	U-L-	This work
JMY8027	Po1d + URA3-scARO4K229L-scARO7T226I + NAT-AtPAL-AtC4H-AtCPR + HPH-AtCHSI-At4CL2-AtCHI1	L-	This work
JMY8028	Po1d + URA3-ARO1-ARO2 + scARO4K229L-scARO7T226I + NAT-AtPAL-AtC4H-AtCPR + HPH-AtCHSI-At4CL2-AtCHI1	L-	This work
JMY8198	Po1d + LEU2-RtTAL-MsCHI-PhCHS	U-	This work
JMY8249	Po1d +LEU2- RtTAL-MsCHI-PhCHS + URA3-At4Cl3 + NAT-PhCHS3	Prototroph	This work
JMY8227	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + LEU2-RtTAL-MsCHI-PhCHS +URA3-At4CL3	Prototroph	This work
JMY8247	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + LEU2-RtTAL-MsCHI-PhCHS +URA3-At4CL3 + NAT-PhCHS3	Prototroph	This work
JMY8294	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + RtTal-MsCHI-PhCHS +At4Cl3 + PhCHS3 + NAT-PhCHS3 + URA3-At4Cl3	L-	This work
JMY8290	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + RtTal-MsCHI-PhCHS +At4Cl3 + PhCHS3 + URA3-ACC1(S659A, S1157A) + LEU2-RtTal-MsCHI-PhCHS	Prototroph	This work
JMY8293	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + RtTal-MsCHI-PhCHS +At4Cl3 + PhCHS3 + URA3-ACC1(S659A, S1157A) + atPAL-atC4H-CPR_Nat	L-	This work
JMY8200	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + NAT-scARO3K222L + LEU2-RtTAL-MsCHI-PhCHS	U-	This work
JMY8284	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + scARO3K222L + RtTAL-MsCHI-PhCHS + URA3-At4CL3 + NAT-PhCHS + LEU2-At4CL3-At4CL2	Prot	This work
JMY8286	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + scARO3K222L + RtTAL-MsCHI-PhCHS + URA3-At4CL3 + NAT-PhCHS +LEU2-RtTAL-MsCHI-PhCHS	Prototroph	This work
JMY8225	Po1d + ARO1-ARO2-ARO4K221L-ARO7G139S + NAT-scARO3K222L + LEU2-RtTAL-MsCHI-PhCHS + URA3-At4CL3	Prototroph	This work
JMY8263	Po1d + ARO1-ARO2-ARO4K221L-ARO7G139S + scARO3K222L + RtTal-MsCHI-PhCHS + URA3-At4Cl3 + NAT-PhCHS	L-	This work
	Resveratrol		
JMY7969	Po1d + HPH-HaTAL-At4CL2-VvSTS	U-L-	This work
JMY8190	Po1d + LEU2-HaTAL-At4CL-VvSTS	U-	This work
JMY8203	Po1d + HPH-HaTAL-At4CL2-VvSTS + LEU2-At4CL3-AtCL2	U-	This work
JMY8213	Po1d + LEU2-HaTAL-At4CL-VvSTS + URA3-At4CL3	Prototroph	This work
JMY7973	Po1d + URA3-scARO4K229L-scARO7T226I + HPH-HaTAL-At4CL2-VvSTS	L-	This work

JMY8264	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + haTAL-at4CL2-vvSTS + URA3-ACC(S659A, S1157A)	L-	This work
JMY8304	Y8264 + At4CL3-AtCL2	Prototroph	This work
JMY8196	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + LEU2-HaTAL-At4CL-VvSTS	U-	This work
JMY8215	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + LEU2-HaTAL-At4CL-VvSTS + URA3-At4CL3	Prototroph	This work
JMY8192	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + NAT-scARO3K222L +LEU2-HaTal-At4Cl-VvSTS	U-	This work
JMY8214	Po1d + ARO1-ARO2 + scARO4K229L-scARO7T226I + NAT-scARO3K222L + LEU2-HaTAL-At4CL-VvSTS + URA3-At4CL3	Prototroph	This work
JMY8194	Po1d + ARO1-ARO2-ARO4K221L-ARO7G139S + NAT-scARO3K222L + LEU2-HaTal-At4Cl-VvSTS	U-	This work
	Melanin		
JMY7735	Po1d + URA3-ARO1-ARO2 + LEU2-scARO4K229L-scARO7T226I	Prototroph	This work
JMY7997	Po1d + LEU2-ARO1-ARO2 + URA3-ARO4K221L-ARO7G139S + HPH-scARO3K222L + NAT-ARO8-ARO10	Prototroph	This work
JMY8131	JMY8032 Δ-*4hppd*	Prototroph	This work
JMY8178	JMY8032 + HPH-4HPPD	Prototroph	This work
JMY8208	JMY8032 + (HPH-4HPPD)x3_SuperBrown Strain	Prototroph	This work
JMY8207	JMY7735 + HPH-4HPPD	Prototroph	This work
	Strains for AAA biosensor tests		
Plate1-B2	JMY2101 + URA-PrAro8-RedStarII-Tlip	Prototroph	This work
Plate1-C2	JMY2101 + URA-PrAro9-RedStarII-Tlip	Prototroph	This work
Plate1-E2	JMY2101 + URA-PrAro10-RedStarII-Tlip	Prototroph	This work
JMY7739	Po1d + VIOA-VIOB-VIOE; ura3-302 leu2-270 xpr2-322	L- U-	This work
JMY7751	Y7739 + URA3-ARO1-ARO2	L-	This work
JMY7755	Y7739 + LEU2-scARO4K229L-scARO7T226I	U-	This work
JMY7759	Y7739 + URA3-scARO4K229L-scARO3K222L	L-	This work
JMY7763	Y7739 + URA3-ARO1-ARO2 + LEU2-scARO4K229L-scARO7T226I	Prototroph	This work
JMY7793	Y7739 + URA3-ARO1ARO2 + LEU2-scARO4K229L-scARO7T226I	Prototroph	This work
JMY7795	Y7739 + URA3-ARO1ARO2 + LEU2-scARO4K229L-scARO3K222L	Prototroph	This work

6.1.4 **Table of primers**

The most relevant primers used are shown here

Name	Sequence	Utilization
Aro1_GGP_G1_Fw	GGGGGTCTCTAATGTTTGCCGAGGGTCAGATC	Amplification of gene from W29
Aro1_GGP_G1_Rv	CCCGGTCTCTTAGAACAGTACATA-CACGTTCAGTCAATACACATTTTAG	Amplification of gene from W29
Aro1_Oligo 1 (Insert 1) Fw	TGGAATTCGCCCTTGCGGCCGGGGG-TCTCTAATGTTTGCC	For Point Mutation
Aro1_Oligo 2 (Insert 1) Rv	TGTGCTCCTTAGTCTCGTCCTTGGTGAC	For Point Mutation
Aro1_Oligo 3 (Insert 2) Fw	GGACGAGACTAAGGAGCACACCTACCACATTCC	For Point Mutation
Aro1_Oligo 4 (Insert 2) Rv	CGGTCTTGAGGCCCAGGTTGACCAGGTTG	For Point Mutation
Aro1_Oligo 5 (Insert 3) Fw	CAACCTGGGCCTCAAGACCGCCGTTGAGC	For Point Mutation
Aro1_Oligo 6 (Insert 3) Rv	AGATGCATGCTCGAGCGGCCCCCGGTCTCTTA-GAACAGTACATAC	For Point Mutation
Aro2_GGP_G2_Fw	GGGGGTCTCTACAAAT-GAGCACTTTCGGCACGCTTTTC	Amplification of gene from W29
Aro2_GGP_G2_Rv	CCCGGTCTCTATCCACA-CATTGAATGCAGTCTGTTGGC	Amplification of gene from W29
Aro2_Oligo 1 (Insert 1) Fw	TGGAATTCGCCCTTGCGGCCGGGGGTCTCTA-CAAATGAGC	For Point Mutation
Aro2_Oligo 2 (Insert 1) Rv	GCTCGCCGAGGCCGACGGGGCAGTTCCG	For Point Mutation
Aro2_Oligo 3 (Insert 2) Fw	CCCCGTCGGCCTCGGCGAGCCCTGTTTC	For Point Mutation
Aro2_Oligo 4 (Insert 2) Rv	AACTTTTCTGATCTCTCAGCTACTGTACTCCCTCC	For Point Mutation
Aro2_Oligo 5 (Insert 3) Fw	GCTGAGAGATCAGAAAAGTTCATGGTGCAGCG	For Point Mutation
Aro2_Oligo 6 (Insert 3) Rv	AGATGCATGCTCGAGCGGCCCCCGGTCTC-TATCCACACATTG	For Point Mutation
Aro4_GGP_G1_Fw	GGGGG-TCTCTAATGTCCCGTTCCTCCTCTCCCAACG	Amplification of gene from W29
Aro4_GGP_G1_Rv	CCCGGTCTCTTAGACAACTCACTATACTTTTA-CATTTTACTATTCACCTG	Amplification of gene from W29
Aro4_Oligo 1 (Insert 1) Fw	TGGAATTCGCCCTTGCGGCCGGGGG-TCTCTAATGTCCCG	For Point Mutation
Aro4_Oligo 2 (Insert 1) Rv	CCTTGGAGACAGTCTCTCGGGCGTTTTTGG	For Point Mutation
Aro4_Oligo 3 (Insert 2) Fw	CCGAGAGACTGTCTCCAAGGGCCGAAAG	For Point Mutation
Aro4_K231L_Oligo4(insert2) Rv	TTACCCAGGGTGGTGGTGATGGCGGC	For Point Mutation
Aro4_K231L_ Oligo5(in-sert3) Fw	CACCACCCTGGGTAACGAGAACTGCTTCATCA	For Point Mutation
Aro4_Oligo 6 (Insert 3) Rv	TGTCCACGGTAGTCTCCCAAGAGACACAGGC	For Point Mutation
Aro4_Oligo 7 (Insert 4) Fw	TTGGGAGACTACCGTGGACATGCTCACCG	For Point Mutation
Aro4_Oligo 8 (Insert 4) Rv	AGATGCATGCTCGAGCGGCCCCCGGTCTCTTA-GACAACTCAC	For Point Mutation

Aro7_GGP_G2_Fw	GGGGGTCTCTACAAATGGACTTCAC-TAAAGCCGACACC	Amplification of gene from W29
Aro7_GGP_G2_Rv	GGTCTCTATCCCCATACCGTTTTTCA-CATTCAATGTAC	Amplification of gene from W29
Aro7_Oligo 1 (Insert 1) Fw	TTCGCCCTTGCGGCCGGGGGTCTCTA-CAAATGGAC	For Point Mutation
Aro7_G139S_Oligo2(insert1) Rv	ACCGAGGAATAGTTCTCGGGCTGATCTCCAG	For Point Mutation
Aro7_G139S_Oligo3(insert2) Fw	GAACTATTCCTCGGTCATGGTGTGCGAC	For Point Mutation
Aro7_Oligo 4 (Insert 2) Rv	CATGCTCGAGCGGCCGGTCTCTATCCCCATACCG	For Point Mutation
Aro9pr_GGP_P1_fw	GGGGGTCTCTACGGTGGAACGGTCA-GAATCGCAATATCCTTC	Promoter region amplification from W29
Aro9pr_GGP_P1_Rv	cccGGTCTCtCATTGCTGGAAACAGTGATGATA-TAGTTGGTGTAACTGT	Promoter region amplification from W29
Aro8pr_GGP_P1_fw	GGGGGTCTCTACGGCAAGTGCGAGCCGTT-GAAGATTG	Promoter region amplification from W29
Aro8pr_GGP_P1_Rv	cccGGTCTCtCATTCTTTGTAGTCGTGCTTTTGG-TGTTGTACGAG	Promoter region amplification from W29
Aro10pr_GGP_P1_fw	GTATGTACGATAGTCGTCTGATCAGGTG	Promoter region amplification from W29
Aro10pr_GGP_P1_Rv	cccGGTCTCtCATTGTTGATAGTCAAGTCACTGGA-GAGATGGT	Promoter region amplification from W29
Aro1-qRTPCR_Fw	ACTCGCGGTTCAAGACTGAT	qRT-PCR
Aro1-qRTPCR_Rv	CAGGAATGATGGTGTTGACG	qRT-PCR
Aro2-qRTPCR_Fw	CAAGATTGTGGAGGAGCACA	qRT-PCR
Aro2-qRTPCR_Rv	GCAAAGCCAGAACCAAACTC	qRT-PCR
Aro3-qRTPCR_Fw	AAACGAGGACTGCTTCCTGA	qRT-PCR
Aro3-qRTPCR_Rv	CAATCTGCTGAGCCACGTTA	qRT-PCR
Aro4-qRTPCR_Fw	TTGGTTTCAAGAACGGAACC	qRT-PCR
Aro4-qRTPCR_Rv	CTCGCAGAATGATGAAGCAG	qRT-PCR
Aro7-qRTPCR_Fw	CCTACAGCAAGGAGGTGAGC	qRT-PCR
Aro7-qRTPCR_Rv	AACTTGCCAAAGTGGATTCG	qRT-PCR
Aro8_qRTPCR_Fw	CAGATGGACGAGTTCAAGCA	qRT-PCR
Aro8_qRTPCR_Rv	AGGAGCCAGAACCTTGGAGT	qRT-PCR
Aro9_qRTPCR_Fw	TCGCAACAGAAGACACCAAC	qRT-PCR
Aro9_qRTPCR_Rv	CTTCGACTTGGGTGGATTGT	qRT-PCR
ACT-qRTPCR_Fw	TCCAGGCCGTCCTCTCCC	qRT-PCR
ACT-qRTPCR_Rv	GGCCAGCCATATCGAGTCGCA	qRT-PCR
RtTAL_qRTPCR_Fw	CAATTTGCGTGATGAATTGG	qRT-PCR
RtTAL_qRTPCR_Rv	CACAGCAGCCAAAGACAAAG	qRT-PCR
MsCHI_qRTPCR_Fw	GGTATCCTCGGCTTGTCCTT	qRT-PCR

MsCHI_qRTPCR_Rv	CAAGCACCGTTTCAAGTCG	qRT-PCR
At4Cl3_qRTPCR_Fw	ATGAGGAAGGATGGTTGCAT	qRT-PCR
At4CL3_qRTPCR_Rv	GGCGTCAGCGATAGAGTGAT	qRT-PCR
PhCHS_qRTPCR_Fw	CCTGCAATTCTCGACCAAGT	qRT-PCR
PhCHS_qRTPCR_Rv	CTGTAGTGCCCAGACCTTCC	qRT-PCR
Aro3scK222L_qRTPCR_Fw	CCCAAAGTGGCTAAGTGCAT	qRT-PCR
Aro3scK222L_qRTPCR_Rv	CACTGAGCAGCCGTATTTCA	qRT-PCR
Aro4scK229L_qRTPCR_Fw	CTGCGAACAGATAGCGAATG	qRT-PCR
Aro4scK229L_qRTPCR_Rv	CAGAACATCCTCGGTGGTTT	qRT-PCR
Aro7scT226I_qRTPCR_Fw	AATTCGTCGCAGAGGCTAAA	qRT-PCR
Aro7scT226I_qRTPCR_Rv	TGGGATCAACTCCGTAAACC	qRT-PCR
HaTAL_qRTPCR_Fw	CCAACAATACATCGCCATTG	qRT-PCR
HaTAL_qRTPCR_Rv	GGTCGCTCGTGAGAGATAGG	qRT-PCR
At4Cl2_qRTPCR_Fw	GCTGGAGTCGCTTCTTATCG	qRT-PCR
At4Cl2_qRTPCR_Rv	CTTCGACACGAACTGCTTGA	qRT-PCR
VvSTS_qRTPCR_Fw	GGCAACCTGAGAGAAGTTGG	qRT-PCR
VvSTS_qRTPCR_Rv	GGCGATCCAGAACAGTGAGT	qRT-PCR
Aro10_qRTPCR_Fw	AGCTGCTCTAGCCAAACGAG	qRT-PCR
Aro10_qRTPCR_Rv	CATCGTTGTTGAGCAGGAAA	qRT-PCR

6.2   **Supplementary articles**

The articles introduced during the manuscript in which I contributed and were part of the development of this work, but that were not included in the main text, are presented hereafter. The order in which they were cited in the text is conserved here.

Melendez, H.G., Larroude, M., Park, Y.K., Trebul, P., Nicaud, J.-M., Ledesma-Amaro, R., 2018. Synthetic biology to improve the production of lipases and esterases, Methods in Molecular Biology. https://doi.org/10.1007/978-1-4939-8672-9_13

Celińska, E., Ledesma-Amaro, R., Larroude, M., Rossignol, T., Pauthenier, C., Nicaud, J.-M., 2017. Golden Gate Assembly system dedicated to complex pathway manipulation in Yarrowia lipolytica. Microb. Biotechnol. 10. https://doi.org/10.1111/1751-7915.12605

Park, Y.K., Korpys, P., Kubiak, M., Celinska, E., Soudier, P., Trébulle, P., Larroude, M., Rossignol, T., Nicaud, J.M., 2019. Engineering the architecture of erythritol-inducible promoters for regulated and enhanced gene expression in Yarrowia lipolytica. FEMS Yeast Res. 19. https://doi.org/10.1093/femsyr/foy105

Soudier, P., Larroude, M., Celińska, E., Rossignol, T., Nicaud, J.-M., 2019. Selection of Heterologous Protein-Producing Strains in Yarrowia lipolytica, in: Methods in Molecular Biology (Clifton, N.J.). pp. 153–168. https://doi.org/10.1007/978-1-4939-9024-5_6

Chapter 13

Synthetic Biology to Improve the Production of Lipases and Esterases (Review)

Heber Gamboa-Melendez, Macarena Larroude, Young Kyoung Park, Pauline Trebul, Jean-Marc Nicaud, and Rodrigo Ledesma-Amaro

Abstract

Synthetic biology is an emergent field of research whose aim is to make biology an engineering discipline, thus permitting to design, control, and standardize biological processes. Synthetic biology is therefore expected to boost the development of biotechnological processes such as protein production and enzyme engineering, which can be significantly relevant for lipases and esterases.

Key words Synthetic biology, Lipases, Esterases, Metabolic engineering, Computer-aided design (CAD), Flux balance analysis (FBA), Genome-scale modelling (GEM), CRISPR-Cas9, ZFN, TALEN

1 Introduction

Lipases and esterases are of main importance for the metabolism of a large number of compounds including fat triacylglycerols as well as xenobiotics, drugs, and environmental pollutants [1]. Differentiated on the basis of their substrate specificity and distribution within organisms and tissues, these enzymes are of great interest in many biotechnological applications, such as the production of structured lipids for the food industry, the production of biodiesel, or the synthesis of polyesters [2].Thus, efforts on producing them in a more cost and time efficient way as well as with improved catalytic properties are of strong interest.

Synthetic biology is an emergent field of research whose aim is to make biology an engineering discipline, thus permitting to design, control, and standardize biological processes. Synthetic biology is therefore expected to boost the development of biotechnological processes such as protein production and enzyme engineering, which can be significantly relevant for lipases and esterases. At the moment, this discipline is involved in developing novel tools (DNA assembly and editing, high-throughput screening and

microbial biotechnology

Brief report

Golden Gate Assembly system dedicated to complex pathway manipulation in *Yarrowia lipolytica*

Ewelina Celińska,[1,†] Rodrigo Ledesma-Amaro,[2,*,†] Macarena Larroude,[2] Tristan Rossignol,[2] Cyrille Pauthenier[3] and Jean-Marc Nicaud[2,*]

[1]*Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, ul. Wojska Polskiego 48, 60-627 Poznan, Poland.*

[2]*Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France.*

[3]*Institute of System and Synthetic Biology, Universite d'Evry vald'Essonnes, Bt. Geneavenir 6 Genopole Campus 1, 5 rue Henry Desbrueres 91000, Evry, France.*

Summary

In this study, we have adopted Golden Gate modular cloning strategy to develop a robust and versatile DNA assembly platform for the nonconventional yeast *Yarrowia lipolytica*. To this end, a broad set of destination vectors and interchangeable building blocks have been constructed. The DNA modules were assembled on a scaffold of predesigned 4 nt overhangs covering three transcription units (each bearing promoter, gene and terminator), selection marker gene and genomic integration targeting sequences, constituting altogether thirteen elements. Previously validated DNA modules (regulatory elements and selection markers) were adopted as the Golden Gate bricks. The system's operability was demonstrated based on synthetic pathway of carotenoid production. This technology greatly enriches a molecular biology toolbox dedicated to this industrially relevant microorganism enabling fast combinatorial cloning of complex synthetic pathways.

Introduction

Synthetic biology, and in particular combinatorial cloning, derives from engineering concepts of standardization, modularity of building blocks and simplification of assembly lines. Traditional protocols employing restriction digestion and one-by-one element cloning are both time- and cost-inefficient (Celińska and Grajek, 2013; Matthaus *et al.*, 2014). Due to the recent development of DNA assembly techniques for metabolic pathway engineering, a great worldwide effort is now being pursued towards establishing such cloning platforms for an individual organism of interest. In the last years, several DNA modular assembly platforms have been developed and new standards have been defined (Sands and Brent, 2001). Golden Gate (GG) modular cloning system, relying on type IIs restriction enzymes, appears as one of the most robust techniques within this field (Engler *et al.*, 2008; Gao *et al.*, 2013). The core of GG strategy lies in establishing a library of standardized and interchangeable DNA parts, which can be subsequently assembled in a single-step, one-pot reaction. Examples of such GG platforms have been recently reported for *Escherichia coli*, yeast or plant species (Engler *et al.*, 2014; Terfrüchte *et al.*, 2014; Agmon *et al.*, 2015; Kakui *et al.*, 2015; Lee *et al.*, 2015; Mitchell *et al.*, 2015; Iverson *et al.*, 2016; Moore *et al.*, 2016). Nevertheless, many of valuable biotech workhorses are still queuing the line for a customized GG platform, amongst these the non-conventional yeast *Y. lipolytica*, which is a well-established biotechnological chassis for the production of numerous valuable bioproducts (Nicaud, 2012; Ledesma-Amaro *et al.*, 2015; Madzak, 2015; Ledesma-Amaro and Nicaud, 2016). Establishing a GG combinatorial cloning platform for *Y. lipolytica* appears as an urgent need, enabling more time- and cost-efficient assembly of complex DNA constructions and standardization of DNA modules, which can be easily exchanged between different laboratories.

Here, we present a robust modular GG system for expression of one, two or three customizable transcription units (TU) in a versatile cassette for *Y. lipolytica*.

Results and discussion

Defining standards and modular parts of the GG system for Y. lipolytica

The ultimate goal of this study was the development of extensive and versatile system of building blocks (Golden Gate Fragments; GGFs) easily fitted into the general scaffold of GGA (Golden Gate Assembly), enabling fast construction of complex expression vectors dedicated to metabolic engineering of *Y. lipolytica*. To complete this task, a set of thirteen 4 nt overhangs was developed together with the corresponding destination vector system (Fig. 1, Supporting Information) to cover three TUs (all composed of promoter – P; ORF = gene – G; terminator – T; all accompanied with a suffix indicating position of a TU); selection marker gene – M; and integration targeting sites – InsUP and InsDOWN.

Incompatibility of the 4 nt sequence was a prerequisite to avoid the risk of unintended matching of the building blocks in uncontrolled position and/or orientation. Such a system of the overhangs was developed and organized in a general scaffold (Fig. 1). Due to extensive character of the envisioned expression cassettes, each of the overhangs was assigned a letter, to enable unambiguous referencing to the respective position in the GGA. Moreover, the overhangs flanking individual genes (in position G1, G2 or G3; see Fig. 1) were designed to
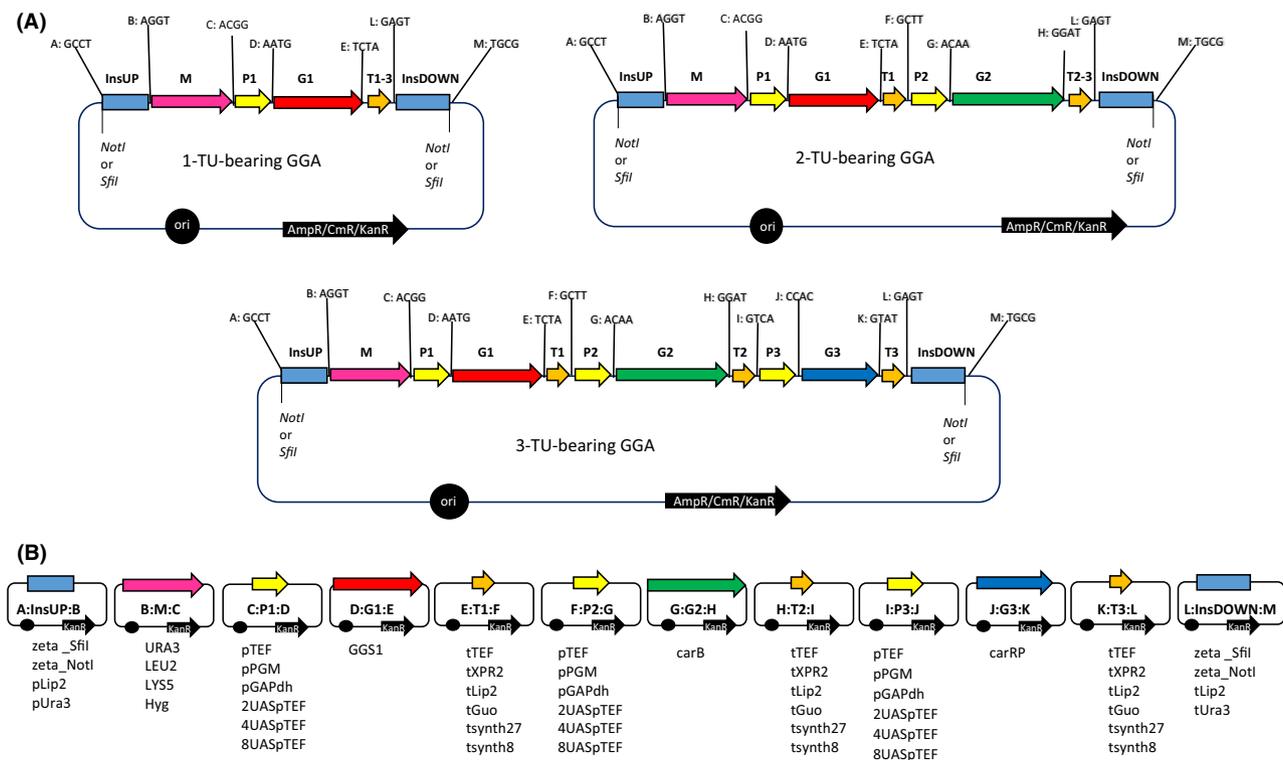

Fig. 1. Golden Gate Assembly platform dedicated to *Yarrowia lipolytica* in a one-, two- and three-transcription unit (TU)-bearing format.
A. Schematic representation of complete GG destination vectors in a one-, two- and three-TU-bearing format. Each of GGA covers at least one TU (composed of promoter – P; gene as ORF – G; and terminator – T) and selection marker (M), which are flanked with integration targeting sequences (InsUP and InsDOWN), constructed on a destination vector backbone, either pSB1A/C/K3-RFP/amilCP/amilGFP for ampicillin (A), chloramphenicol (C) or kanamycin (K) resistance marker respectively; pSB vector series offers exploitation of one of the following reporter genes in *E. coli*: RFP – red chromophore; amilCP – blue chromophore; amilGFP – green fluorophore. Building blocks are flanked with pre-designed 4 nt overhang, each assigned a letter from A–M. Sequences of the overhangs are provided next to the respective position. Abbreviations: InsUP/DOWN – sequence targeting insertion in the 5'/3' region; M – selection marker; P1/2/3 – promoter governing expression of the gene of interest in transcription unit 1, 2 and 3 respectively; G1/2/3 – gene of interest in transcription unit 1, 2 and 3 respectively; T1/2/3 – terminator located in the transcription unit 1, 2 and 3 respectively. NotI/SfiI: NotI/SfiI recognition sites for release of the expression cassettes prior to *Y. lipolytica* transformation. B. Donor vectors constructed on a backbone of pCR II-Blunt TOPO vector (Invitrogen/Life Technologies) bearing kanamycin resistance gene. Each building block is flanked with respective BsaI recognition sites and predesigned 4 nt overhangs. Exemplary building blocks (Golden Gate Fragments; GGFs) constructed in this study are provided as a list underneath the donor vectors' schemes. For reference to nucleotide sequences, see Supporting Information.

minimize the scar between a coding sequence and the regulatory elements. Thus, the overhang prior to the coding sequence was designed to maximally fit into the consensus sequence preceding significant fraction of ORFs in *Y. lipolytica* genome – CACA – or include the start codon (Gasmi *et al.*, 2011).

The GG modular system was designed to include two flanking insertion sites, a selection marker and one, two or three TUs in each assembly. The number of TUs can be easily manipulated by amplification of 'hybrid' terminators, provided that the same sequence was used in the merged positions (T1, T2, T3; compare Fig. 1). These 'hybrid' terminators should be amplified with forward primer with 'E' overhang and reverse primer with 'L' overhang to render T1-3 'hybrid' terminator, skipping the second and third transcription units, or with 'H' overhang bearing forward primer and 'L' overhang bearing reverse primer, to gain T2-3 'hybrid' terminator, skipping the third transcription unit.

The destination vector backbones were equipped with a chromophore (RFP or amilCP) or fluorophore (amilGFP) gene flanked with BsaI sites with predesigned overhangs to define a colour-based negative cloning selection marker. The BsaI recognition sites were positioned outwardly with respect to the backbone of the destination vector, so that after digestion with BsaI, both the reporter gene and the BsaI recognition sites were released, leaving the vector's backbone flanked with protruding 4 nt overhangs. A set of different antibiotic resistance genes was incorporated into the system of the destination vectors, to enable greater flexibility of experimental set-up. In the proof-of-concept experiments (with three TUs bearing cassette; described below), we have used a pSB1K3-RFP variant from the iGEM collection (http://parts.igem.org/Collections).

It is commonly accepted that only integrative expression cassettes constitute truly potent tools in *Y. lipolytica* genetic engineering, as the so far developed episomal plasmids tend to be unstable and low-copy (Liu *et al.*, 2014). This fact ruled the construction of both the destination plasmids as well as the scaffold of GGA, flanked with GGFs targeting integration of the expression cassette into the host genome (InsUP and InsDOWN). Thus, the destination vector's backbone containing solely bacterial elements can be discarded prior to transformation of *Y. lipolytica* cells. To enable the release of the expression cassette composed of transcription units, selection marker and targeting sequences, from the bacterial backbone, rare-cutting restriction endonuclease recognition sites NotI or SfiI, depending on the targeted GGA structure, were added at the borders of the expression cassette in the InsUP and InsDOWN fragments (Fig. 1).

Generation of GG library dedicated to Y. lipolytica

All the sequences to be used as building blocks of the envisioned GGA were extracted from *Y. lipolytica* W29 genome sequence or one of the previously constructed vectors from our own collection. All the sequences were analysed *in silico* to find internal BsaI site, to be eventually eliminated using assembly PCR technique (Sambrook and Russell, 2001). Subsequently, an extensive set of primers (referred to as GGP, for Golden Gate Primers) equipped with the predesigned system of the 4 nt overhangs and externally located BsaI recognition sites was designed (Supporting Information; Pauthenier and Faulon, 2014). The set of GGF presented in this work, the interchangeable parts for promoter, terminator, selection marker and integration site (Fig. 1B), constitutes the response to our current needs, but can be easily expanded on the other building blocks of interest by simply adding the required prefixes and suffixes within the primers at the stage of the element amplification. As presented in Fig. 1, the major focus of this study was to develop a system of easily tractable phenotype of positive transformants. Therefore, for the genes of interest we have chosen the complete pathway for carotenoid production, for convenient traceability of positive clones. With the same concept in mind, *LIP2* and *URA3* loci were selected as the targeting platforms, and thus, correct integration could be verified on tributyrin- or sucrose-containing plate respectively. A *LIP2* deletion results in a reduced halo of triglyceride hydrolysis (Pignède *et al.*, 2000), while replacement of *ura3-320* allele bearing *S. cerevisiae SUC2* gene gives transformants unable to grow on sucrose (Nicaud *et al.*, 1989; Barth and Gaillardin, 1996). Zeta sequences were chosen to reach random integrations in Po1d strain (JMY195) or at a zeta docking platform in Po1d derivative JMY1212, bearing a zeta sequence at the *leu2-270* locus (Bordes *et al.*, 2007).

All the amplicons bearing respective GGFs were cloned in the donor vectors (pCR Blunt II TOPO vectors; Thermo Fisher Scientific, Villebon sur Yvette, France), verified by BsaI digestions and sequencing and deposited in our collection. The structure of the targeted GGA cassette ruled the composition of the GG reaction mixture. The Golden Gate reaction conditions were designed based on the previously published protocols (Engler *et al.*, 2008; Pauthenier *et al.*, 2012; Agmon *et al.*, 2015). The reaction mixture contained precalculated equimolar amount of each GGF and the destination vector (50 pmoles of ends), 2 μl of T4 DNA ligase buffer (NEB), 5 U of BsaI, 200 U of T4 and ddH2O up to 20 μl. The following thermal profile was applied: [37°C for 5 min, 16°C for 2 min] × 60, 55°C for 5 min, 80°C for 5 min, 15°C ∞. Subsequently, the reaction mixture was

used for *E. coli* DH5α transformation (Sambrook and Russell, 2001). White colonies were screened for identification of complete GGA through plasmid isolation, restriction digestion and multiplex PCR. Complete GGA was subsequently linearized and used for transformation of either *Y. lipolytica* Po1d *ura- leu-* (JMY195 *MatA leu2-270 ura3-302 xpr2-322*) or Po1d *ura- leu+* bearing a zeta docking platform (JMY1212; *MatA ura3-302 xpr2-322, LEU2, zeta*; Bordes *et al.*, 2007) via the lithium acetate transformation protocol (Chen *et al.*, 1997). Derivative strain Po1d *leu+ ura+* (JMY2900; *MatA ura3-302 xpr2-322*) was used as prototroph control. Clones bearing complete GGA integrated with the host genome exhibited red–orange phenotype due to carotenoid production. All the GGF building blocks included in the GGA have been proven functional in previous experiments (see Table S1).

Proof of concept based on carotenoid-pathway-containing Golden Gate Assembly

To validate robustness and efficiency of the here developed GG system, we have constructed a complex assembly structure comprising twelve GGFs and the destination vector, covering three TUs, accompanied by selection marker and integration targeting sequences (Fig. 2A).

The three genes within this assembly were native GGS1 (geranylgeranyl diphosphate synthase; YALI0D17050g) in G1 position and heterologous carB (phytoene dehydrogenase; AJ238028.1) and carRP (phytoene synthase; AJ250827.1) coding sequences from *Mucor circinelloides* in G2 and G3 positions respectively (sequences were kindly provided by DSM

Company). Such a set of genes was previously proved to be efficiently expressed in *Y. lipolytica* cells, constructed using DNA assembler strategy (Gao *et al.*, 2014) or classical cloning techniques (Matthaus *et al.*, 2014). All the genes were assembled with corresponding regulatory elements, zeta insertion sites and *URA3* selection marker. The correct complete GGA was first preselected in *E. coli* clones through multiplex PCR and restriction digestion and then linearized and transformed into competent *Y. lipolytica* JMY195 cells. On average, more than 4×10^2 colonies could be obtained from a single transformation reaction. About 10% of the transformants were white (exemplary results for one transformation run: 50/517, 91/1052, 41/396). The remaining clones exhibited different intensity of the orange colour development (Fig. 2B). The observed variability in the carotenoid-producer phenotype could result from varying number of either complete or partial GGA copies integrated within the host genome as observed by Gao *et al.* (2014), although this would be unlikely with the amount of DNA we used (Bordes *et al.*, 2007). More likely, the random distribution of the cassette within the genomic DNA due to the zeta targeting sequences, as investigated in Pignede *et al.* (2000), may be responsible for the colour variability. This hypothesis is supported by the results obtained when the same GGA was used to transform strain JMY1212, bearing a zeta docking platform at the leu2-270 locus enabling homologous recombination. Although the number of white colonies after transformation of this strain was higher (33%; 33/98), the orange colonies did not present the colour variability that was found amongst JMY195-based transformants. In a previous study with DNA assembler strategy, when the same carotenoid-pathway genes were used, the efficiency of
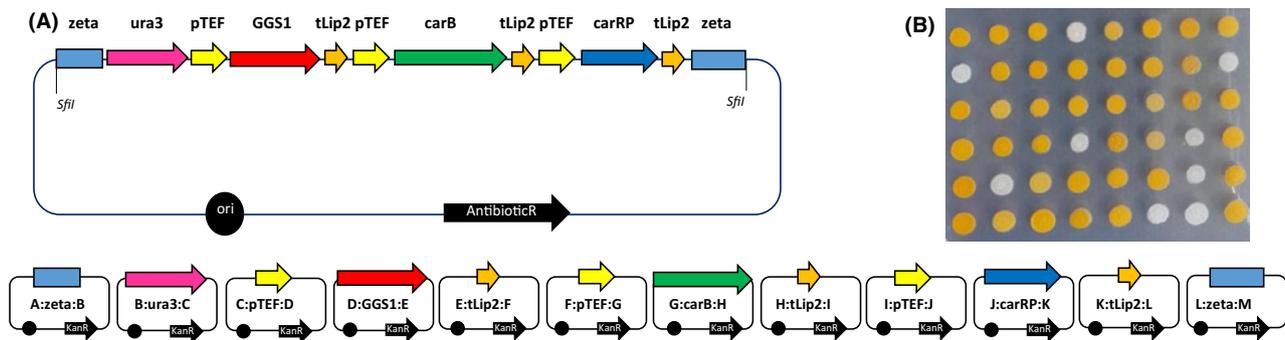

Fig. 2. Golden Gate Assembly bearing 'carotenoid synthesis' pathway dedicated to integration with *Yarrowia lipolytica* genome.
A. Scaffold of GGA comprising three transcription units and selection marker, flanked with integration targeting sequences, constructed on a destination vector backbone. Each gene was flanked with 396 nt of TEF promoter and 122 nt of Lip2 terminator sequences, both native to *Y. lipolytica*. URA3 (1289 nt) gene was used as selection marker in this assembly. Random integrations in Po1d strain (JMY195) were driven through zeta sequences (305 nt and 395 nt for UP and DOWN respectively). Representation of donor vectors bearing respective GGFs is provided below. B. *Y. lipolytica* JMY195 transformants bearing 'carotenoid synthesis' GGA. Variability in colour development is discussed in the Results and Discussion section.

reaching desired phenotype was 20% (Gao *et al.*, 2014). In the current study, the efficiency was largely improved, as 67% to 90% of the obtained *Y. lipolytica* clones exhibited desired red–orange phenotype, for JMY1212 and JMY195 strains respectively. Improved efficiency was accompanied by time- and workload savings, as well as versatility and standardization, being indigenous for modular DNA cloning techniques, altogether making this approach a method of choice for synthetic biology endeavours.

Here, we present a Golden Gate standard specifically designed to modify *Y. lipolytica* in a versatile manner. Robustness and reliability of the system were validated for thirteen-element-bearing GGA, with high efficiency of the desired phenotype recovery in the recombinants. This technology greatly enriches the molecular biology toolbox dedicated to this industrially relevant microorganism, permitting faster and accurate multiple target engineering, a limitless expansion of the building blocks library from endogenous or heterologous origins, or the use of interchangeable modules in combinatorial approaches followed by screening desired phenotypes.

Acknowledgements

Authors' contribution

EC and RLA contributed equally to this work. EC, RLA, ML and TR contributed experimental data. CP participated in the selection of the 4 nt overhangs for the GGA scaffold and provided destination vectors. EC, RLA and JMN wrote the manuscript. The work was performed in the laboratory of TR and JMN.

Conflict of interest

The authors do not have any conflict of interest.

References

Agmon, N., Mitchell, L.A., Cai, Y., Ikushima, S., Chuang, J., Zheng, A., *et al.* (2015) Yeast Golden Gate (yGG) for the efficient assembly of *S. cerevisiae* transcription units. *ACS Synth Biol* **4:** 853–859.

Barth, G., and Gaillardin, C. (1996) *Yarrowia lipolytica*. In *Nonconventional Yeasts in Biotechnology A Handbook*. Wolf, K. (ed). Berlin Heidelberg: Springer, pp. 313–388.

Bordes, F., Fudalej, F., Dossat, V., Nicaud, J.-M., and Marty, A. (2007) A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. *J Microbiol Methods* **70:** 493–502.

Celińska, E., and Grajek, W. (2013) A novel multigene expression construct for modification of glycerol metabolism in *Yarrowia lipolytica*. *Microb Cell Fact* **12:** 102.

Chen, D.C., Beckerich, J.M., and Gaillardin, C. (1997) One-step transformation of the dimorphic yeast *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **48:** 232–235.

Engler, C., Kandzia, R., and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3:** e3647.

Engler, C., Youles, M., Gruetzner, R., Ehnert, T.-M., Werner, S., Jones, J.D.G., *et al.* (2014) A Golden Gate modular cloning toolbox for plants. *ACS Synth Biol* **3:** 839–843.

Gao, X., Yan, P., Shen, W., Li, X., Zhou, P., and Li, Y. (2013) Modular construction of plasmids by parallel assembly of linear vector components. *Anal Biochem* **437:** 172–177.

Gao, S., Han, L., Zhu, L., Ge, M., Yang, S., Jiang, Y., and Chen, D. (2014) One-step integration of multiple genes into the oleaginous yeast *Yarrowia lipolytica*. *Biotechnol Lett* **36:** 2523–2528.

Gasmi, N., Fudalej, F., Kallel, H., and Nicaud, J.-M. (2011) A molecular approach to optimize hIFN α2b expression and secretion in *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **89:** 109–119.

Iverson, S.V., Haddock, T.L., Beal, J., and Densmore, D.M. (2016) CIDAR MoClo: improved MoClo assembly standard and new *E. coli* part library enable rapid combinatorial design for synthetic and traditional biology. *ACS Synth Biol* **5:** 99–103.

Kakui, Y., Sunaga, T., Arai, K., Dodgson, J., Ji, L., Csikász-Nagy, A., *et al.* (2015) Module-based construction of plasmids for chromosomal integration of the fission yeast *Schizosaccharomyces pombe*. *Open Biol* **5:** 150054.

Ledesma-Amaro, R., and Nicaud, J.M. (2016) *Yarrowia lipolytica* as a biotechnological chassis to produce usual and unusual fatty acids. *Prog Lipid Res* **61:** 40–50.

Ledesma-Amaro, R., Dulermo, T., and Nicaud, J.M. (2015) Engineering *Yarrowia lipolytica* to produce biodiesel from raw starch. *Biotechnol Biofuels* **8:** 148.

Lee, M.E., DeLoache, W.C., Cervantes, B., and Dueber, J.E. (2015) A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth Biol* **4:** 975–986.

Liu, L., Otoupal, P., Pan, A., and Alper, H.S. (2014) Increasing expression level and copy number of a *Yarrowia lipolytica* plasmid through regulated centromere function. *FEMS Yeast Res* **14:** 1124–1127.

Madzak, C. (2015) *Yarrowia lipolytica*: recent achievements in heterologous protein expression and pathway engineering. *Appl Microbiol Biotechnol* **99:** 4559–4577.

Matthaus, F., Ketelhot, M., Gatter, M., and Barth, G. (2014) Production of lycopene in the non-carotenoid-producing

yeast *Yarrowia lipolytica*. *Appl Environ Microbiol* **80:** 1660–1669.

Mitchell, L.A., Chuang, J., Agmon, N., Khunsriraksakul, C., Phillips, N.A., Cai, Y., *et al.* (2015) Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. *Nucleic Acids Res* **43:** 6620–6630.

Moore, S.J., Lai, H., Kelwick, R., Chee, S.M., Bell, D., Polizzi, K.M., and Freemont, P.S. (2016) EcoFlex – a multifunctional MoClo kit for *E. coli* synthetic biology. *ACS Synth Biol* **5:** 1059–1069.

Nicaud, J.M. (2012) *Yarrowia lipolytica*. *Yeast* **29:** 409–418.

Nicaud, J., Fabre, E., and Gaillardin, C. (1989) Expression of invertase activity in *Yarrowia lipolytica* and its use as a selective marker. *Curr Genet* **16:** 253–260.

Pauthenier, C., and Faulon, J.L. (2014) PrecisePrimer: an easy-to-use web server for designing PCR primers for DNA library cloning and DNA shuffling. *Nucleic Acids Res* **42:** 205–209.

Pauthenier, C., El-Sayyed, H., Rostain, W., Cerisy, T., Gallo, C., Ferreira, R., *et al.* (2012) BBF RFC 92: The GoldenBricks assembly: A standardized one-shot cloning technique for complete cassette assembly.

Pignede, G., Wang, H.J., Fudalej, F., Seman, M., Gaillardin, C., and Nicaud, J.M. (2000) Autocloning and amplification of LIP2 in *Yarrowia lipolytica*. *Appl Environ Microbiol* **66:** 3283–3289.

Pignède, G., Wang, H., Fudalej, F., Gaillardin, C., Seman, M., and Nicaud, J.M. (2000) Characterization of all extracellular lipase encoded by LIP2 in *Yarrowia lipolytica*. *J Bacteriol* **182:** 2802–2810.

Sambrook, J., and Russell, D. (2001) *Molecular Cloning: A Laboratory Manual*, 3rd edn. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Sands, B., and Brent, R. (2016) Overview of post Cohen-Boyer methods for single segment cloning and for multisegment DNA assembly. *Current Protocols in Molecular Biology* **113:** 3.26.1–3.26.20.

Terfrüchte, M., Joehnk, B., Fajardo-Somera, R., Braus, G.H., Riquelme, M., Schipper, K., and Feldbrügge, M. (2014) Establishing a versatile Golden Gate cloning system for genetic engineering in fungi. *Fungal Genet Biol* **62:** 1–10.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. A list of plasmids and oligonucleotides used in this study.

RESEARCH ARTICLE

Engineering the architecture of erythritol-inducible promoters for regulated and enhanced gene expression in *Yarrowia lipolytica*

Young-Kyoung Park[1], Paulina Korpys[1,2], Monika Kubiak[1,2], Ewelina Celińska[2], Paul Soudier[1], Pauline Trébulle[1], Macarena Larroude[1], Tristan Rossignol[1,†] and Jean-Marc Nicaud[1,*,‡]

[1]Micalis Institute, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France and [2]Department of Biotechnology and Food Microbiology, Poznan University of Life Sciences, ul. Wojska, Polskiego 48, 60-627 Poznan, Poland

*Corresponding author: Micalis Institute, INRA-AgroParisTech, UMR1319, Team BIMLip: Integrative Metabolism of Microbial Lipids, domaine de Vilvert, 78352 Jouy-en-Josas, France. Tel: +33 1 74 07 18 20; jean-marc.nicaud@inra.fr

One sentence summary: This study identified cis-regulatory modules (CRMs) for the *EYK1* and *EYD1* promoters in *Yarrowia lipolytica*, which allowed the development of erythritol-inducible hybrid promoters with practical applications in metabolic engineering and synthetic biology.

†Tristan Rossignol, http://orcid.org/0000-0003-0718-0684

‡Jean-Marc Nicaud, http://orcid.org/0000-0002-6679-972X

ABSTRACT

The non-conventional model yeast *Yarrowia lipolytica* is of increasing interest as a cell factory for producing recombinant proteins or biomolecules with biotechnological or pharmaceutical applications. To further develop the yeast's efficiency and construct inducible promoters, it is crucial to better understand and engineer promoter architecture. Four conserved cis-regulatory modules (CRMs) were identified via phylogenetic footprinting within the promoter regions of *EYD1* and *EYK1*, two genes that have recently been shown to be involved in erythritol catabolism. Using CRM mutagenesis and hybrid promoter construction, we identified four upstream activation sequences (UASs) that are involved in promoter induction by erythritol. Using RedStarII fluorescence as a reporter, the strength of the promoters and the degree of erythritol-based inducibility were determined in two genetic backgrounds: the *EYK1* wild type and the *eyk1Δ* mutant. We successfully developed inducible promoters with variable strengths, which ranged from 0.1 SFU/h to 457.5 SFU/h. Erythritol-based induction increased 2.2 to 32.3 fold in the *EYK1* + wild type and 2.9 to 896.1 fold in the *eyk1Δ* mutant. This set of erythritol-inducible hybrid promoters could allow the modulation and fine-tuning of gene expression levels. These promoters have direct applications in protein production, metabolic engineering and synthetic biology.

Keywords: *Yarrowia lipolytica*; promoter; inducible; erythritol; Golden Gate; gene expression; synthetic biology

INTRODUCTION

Yarrowia lipolytica is an oleaginous yeast species that serves as a non-conventional model organism in research on lipid turnover and bio-oil production (Beopoulos *et al.* 2008, 2009), dimorphic transition and fungal differentiation (Martinez-Vazquez *et al.* 2013), and secretory protein synthesis (Matoba *et al.* 1988; Matoba and Ogrydziak 1989; Boisramé *et al.* 1998; Pignède *et al.* 2000; Nicaud *et al.* 2002). *Y. lipolytica* is also the focus of increasing

Selection of Heterologous Protein-Producing Strains in *Yarrowia lipolytica* ₂ ₃

Paul Soudier, Macarena Larroude, Ewelina Celinska, Tristan Rossignol, ₄
and Jean-Marc Nicaud ₅

Abstract ₆

Yarrowia lipolytica has emerged as an alternative expression system for heterologous protein production ₇ [AU1]
and enzyme evolution. Several different expression systems dedicated for this species have been developed, ₈
ranging from the simple cloning of expression vectors to recently developed high-throughput methodol- ₉
ogies using efficient cloning and assembly such as Gateway and Golden Gate strategies. The latter strategies, ₁₀
due to their modular character, enable multiple vector construction and the construction of expression ₁₁
cassettes containing different genes or a gene under different promoters of various strengths. ₁₂

Here, we present the Golden Gate cloning strategy for the construction of multiple expression cassettes, ₁₃
the transformation into *Y. lipolytica*, and the selection of efficient enzyme-producing strains using an insect ₁₄
alpha-amylase as a reporter detected via thermal cycler-based microassay. ₁₅

Key words *Yarrowia lipolytica*, Heterologous protein, Secretion, Golden Gate, Gateway, Gene ₁₆
assembly, Inducible promoter, Targeting sequence, High-throughput techniques ₁₇

1 Introduction ₁₈

Yarrowia lipolytica has emerged as an attractive alternative host for ₁₉
heterologous protein production, thanks to its ability to efficiently ₂₀
secrete protein due to co-translational translocation of a polypep- ₂₁
tide, performing low level of glycosylation and having a GRAS ₂₂
status. To have a detailed overview of the heterologous protein ₂₃
expression and secretion in *Y. lipolytica*, see the recent reviews of ₂₄
Madzak and Beckerich [1, 2]. We will here briefly summarize the ₂₅
evolution of the current tools for protein production in ₂₆
Y. lipolytica, the methods and their specific application, and the ₂₇
more recent achievements using high-throughput techniques. ₂₈

The expression system employing *Y. lipolytica* was initially ₂₉
based mainly on integrative vectors, since replicative plasmids are ₃₀
maintained in low copy number and tend to be relatively unstable in ₃₁
the cells. The first plasmids were based mainly on vectors carrying ₃₂

6.3   List of contributions and communications

Scientific articles sorted by year

Larroudé, M., Nicaud, J.-M., Rossignol, T. Golden Gate method for *Yarrowia lipolytica* (chapter book under review in the Methods in Molecular Biology book series.)

Larroude, M., Trabelsi, H., Nicaud, J.-M., Rossignol, T., 2020. A set of Yarrowia lipolytica CRISPR/Cas9 vectors for exploiting wild-type strain diversity. Biotechnol. Lett. https://doi.org/10.1007/s10529-020-02805-4

Larroude, M., Park, Y.K., Soudier, P., Kubiak, M., Nicaud, J.M., Rossignol, T., 2019. A modular Golden Gate toolkit for Yarrowia lipolytica synthetic biology. Microb. Biotechnol. https://doi.org/10.1111/1751-7915.13427

Park, Y.K., Korpys, P., Kubiak, M., Celinska, E., Soudier, P., Trébulle, P., Larroude, M., Rossignol, T., Nicaud, J.M., 2019. Engineering the architecture of erythritol-inducible promoters for regulated and enhanced gene expression in Yarrowia lipolytica. FEMS Yeast Res. 19. https://doi.org/10.1093/femsyr/foy105

Soudier, P., Larroude, M., Celińska, E., Rossignol, T., Nicaud, J.-M., 2019. Selection of Heterologous Protein-Producing Strains in Yarrowia lipolytica, in: Methods in Molecular Biology (Clifton, N.J.). pp. 153–168. https://doi.org/10.1007/978-1-4939-9024-5_6

Larroude, M., Rossignol, T., Nicaud, J.-M., Ledesma-Amaro, R., 2018. Synthetic biology tools for engineering Yarrowia lipolytica. Biotechnol. Adv. 36, 2150–2164. https://doi.org/10.1016/J.BIOTECHADV.2018.10.004

Melendez, H.G., Larroude, M., Park, Y.K., Trebul, P., Nicaud, J.-M., Ledesma-Amaro, R., 2018. Synthetic biology to improve the production of lipases and esterases, Methods in Molecular Biology. https://doi.org/10.1007/978-1-4939-8672-9_13

Celińska, E., Ledesma-Amaro, R., Larroude, M., Rossignol, T., Pauthenier, C., Nicaud, J.-M., 2017. Golden Gate Assembly system dedicated to complex pathway manipulation in   Yarrowia   lipolytica.   Microb.   Biotechnol.   10. https://doi.org/10.1111/1751-7915.12605

Larroude, M., Celinska, E., Back, A., Thomas, S., Nicaud, J.-M., Ledesma-Amaro, R., 2017. A synthetic biology approach to transform Yarrowia lipolytica into a competitive biotechnological producer of β-carotene. Biotechnol. Bioeng. https://doi.org/10.1002/bit.26473

Conferences and workshops

- European Summit of Industrial Biotechnology. November 2019, Grazt, Austria. International conference. Poster presentation.

    1. "Synthetic biology tools for non-Saccharomyces yeasts" Macarena Larroude, Arun S. Rajkumar, Jean-Marc Nicaud, Tristan Rossignol and John P. Morrissey.

    2. "Engineering Yeasts to Produce Aromatics" Else-Jasmijn Hassing, Arun S. Rajkumar, Macarena Larroude, Joel Abidemi Akinola, Philip A. de Groot, Mario Beck, Vita Marquenie, Jack T. Pronk, John P. Morrissey and Jean-Marc G. Daran.

    3. "Y. lipolytica as a platform for production of lipids" Macarena Larroude, Jean-Marc Nicaud, Tristan Rossignol

- Physiology of Yeast and Filamentous Fungi conference. June 2019. Milan, Italy. International conference. Poster Presentation. Poster presentation. "Synthetic biology toolkit for engineering the non-conventional yeast Yarrowia lipolytica" Macarena Larroude, Jean-Marc Nicaud, Tristan Rossignol

- DocMicalis Day. PhD student's presentation day at Micalis Institute. May 2019. Jouy-en-josas, France. Internal symposium. Oral presentation. "Development of synthetic biology tools for an industrial yeasts and construction of tailor-made chassis strains" Macarena Larroude, Jean-Marc Nicaud, Tristan Rossignol

- Workshop "Application of CRISPR technologies in microbial biotechnology". Organized by NBV and CHASSY. Technological University of Delft, Delft, Netherlands, February 2019. Assistant.

- Applied Synthetic Biology in Europe. International conference. October 2018. Toulouse, France. Poster presentation. "Synthetic biology toolkit for engineering the non-conventional yeast Yarrowia lipolytica" Macarena Larroude, Jean-Marc Nicaud, Tristan Rossignol

- BioSynSys conference. International conference. Toulouse, France. October 2018. Poster presentation. "Synthetic biology toolkit for engineering the non-

conventional yeast Yarrowia lipolytica" Macarena Larroude, Jean-Marc Nicaud, Tristan Rossignol

- WorkShop "Enzyme-constrained genome-scale metabolic models workshop" Chalmers University, Gothenburg, Sweden. August 2018. Assistant.

- Non-conventional Yeast Conference. International conference. May 2018. Rzeszow, Poland. Poster presentation. "A synthetic biology approach for pathway engineering for the non-conventional yeast Yarrowia lipolytica". Macarena Larroude, Ewelina Celinska, Alexandre Back, Stéphane Thomas, Tristan Rossignol, Rodrigo Ledesma-Amaro, Jean-Marc Nicaud

- DocMicalis Day. PhD student's presentation day at Micalis Institute. July 2018. Jouy-en-josas, France. Internal symposium. Organizing committee.

- BioSynSys conference. International conference. October 2017. Montpellier, France. Oral presentation and poster presentation. "A synthetic biology approach for pathway engineering for the non-conventional yeast Yarrowia lipolytica". Macarena Larroude, Ewelina Celinska, Alexandre Back, Stéphane Thomas, Tristan Rossignol, Rodrigo Ledesma-Amaro, Jean-Marc Nicaud

- Yeast Lipid Conference. International conference. Organizing committee. Paris, France. 2017.

6.4 Students supervision

2018- February to July (6 month). Internship supervision of 2nd year master student, Yolaine Mounier. Developing chassis strains for the production of high value molecules. Introduction to regular molecular biology tools, utilization of high-throughput method of assembly (Golden Gate), transformation methods, *Y. lipolytica* strains construction, genes mutation, gene overexpression, heterologous synthetic pathway expression.

2018- February to March (2 month). Internship supervision of post 2nd year master student, Monika Kubiak. Construction, verification and characterisation of Golden Gate compatible parts for enlarging the developed tool kit for *Y. lipolytica*. Regular molecular biology tools, utilization of high-throughput method of assembly (Golden Gate), transformation methods, construction of *Y. lipolytica* strains, florescence measure and analysis of data.

2017- May to July (2 month) Internship supervision of 1st year master student, Paul Saudier. Characterisation of different promoters in *Y. lipolytica* (natural, hybrid, constitutive and inducible promoters). Introduction to regular molecular biology tools, utilization of high-throughput method of assembly (Golden Gate), transformation methods, construction of *Y. lipolytica* strains florescence measure and analysis of data.

6.5 **Résumé complet en Français**

L'utilisation de microorganismes comme usines cellulaires pour la production de molécules d'intérêt apparaît comme une alternative écologique à la synthèse chimique. Avec l'avènement d'outils peu coûteux de séquençage et de synthèse de l'ADN et de bio-informatique, cette solution est maintenant une réalité. Néanmoins, l'optimisation de l'hôte microbien reste encore un des facteurs limitants pour répondre aux demandes du marché. Ainsi, l'expansion des souches utilisées comme usines cellulaires, la boîte à outils disponible pour l'ingénierie génomique, les méthodes de criblage à haut débit et la gamme de molécules produites sont des axes à développer. Dans ce sens, la construction de souches rationnellement conçues pour avoir un approvisionnement optimisé en molécules qui peuvent servir pour comme précurseurs pour la synthèse d'une large gamme de produits de grande valeur, sont donc d'un intérêt industriel majeur afin de trouver des processus robustes et rentables.

Dans le but de contribuer à augmenter la disponibilité des outils et des souches qui peuvent être utilisés comme usines cellulaires, ce travail propose la levure dite 'non-conventionnelle' *Y. lipolytica* comme cadre pour la construction de châssis pour la production de composés dérivés des acides aminés aromatiques (AAA) qui ont une pléthore d'applications concernant la santé humaine et la nutrition.

La capacité à modifier le métabolisme d'un microorganisme dépend de la disponibilité d'outils de biologie moléculaire, et c'est l'une des limitations dans l'utilisation d'organismes non conventionnels. Ainsi, avec l'objectif d'augmenter l'efficacité lors de l'ingénierie du métabolisme de *Y. lipolytica*, des outils qui peuvent être appliqués de manière systématique et générique pour construire des cassettes d'expression et modifier le génome de *Y. lipolytica* ont été développés au cours de ces travaux. Pour l'assemblage de l'ADN, une technique de clonage modulaire basée sur Golden Gate (GG) (Engler et al., 2008) a été mise en place (Figure 43). Cette méthode permet d'assembler dans un pot multiples fragments d'ADN, facilement interchangeables, en appliquant des cycles de digestion-ligation. Il a été conçu pour permettre l'assemblage de 13 éléments génétiques modulaires (briques), dont trois unités de transcription (TU) (chacun composé d'un promoteur, d'un gène et d'un terminateur), un marqueur sélectif et des séquences pour l'intégration dans le génome (Celińska et al., 2017). La construction de
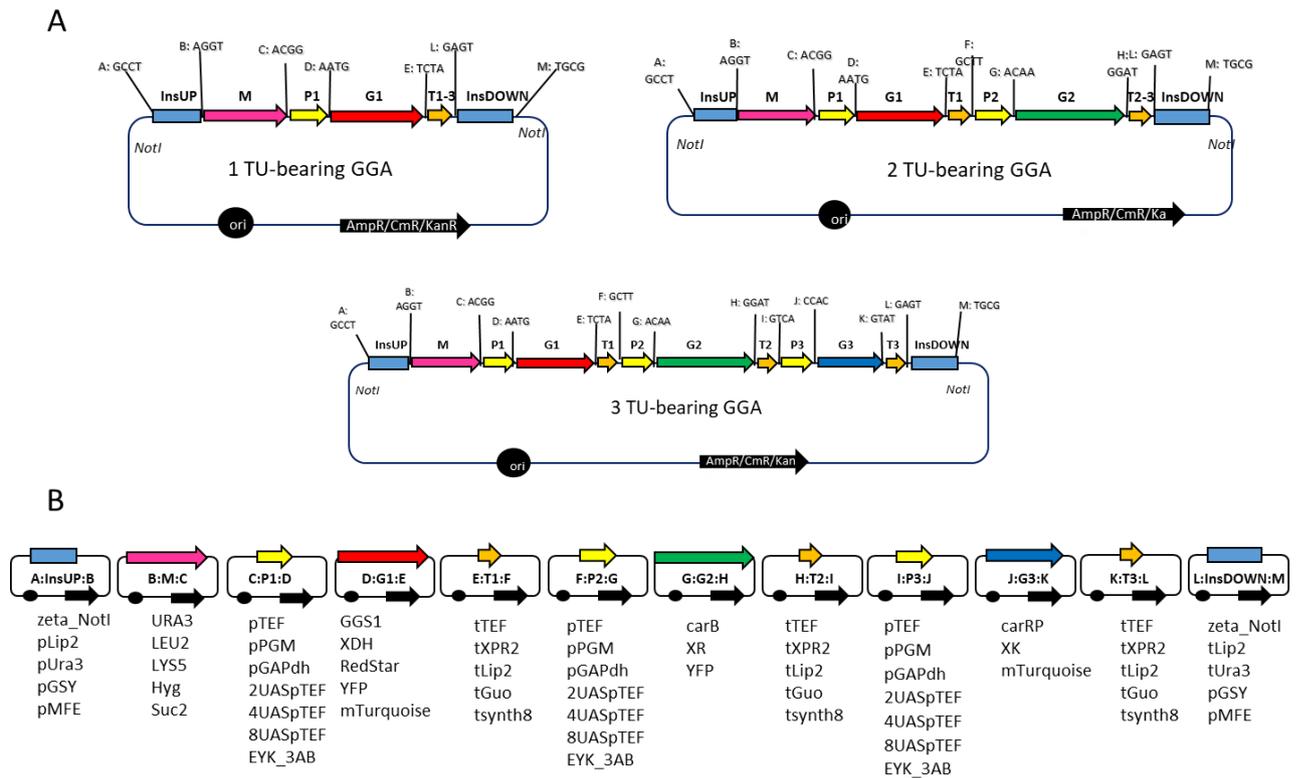
Figure 43. Golden Gate scaffold and library developed for *Y. lipolytica*, allowing the assembly of up to thirteen DNA blocks.

A-Schematic representation of the GG scaffolds that allow the assembly of 1, 2 o 3 TU. **B**- Example of building blocks that are available for each GG position. TU: transcription unit, composed of promoter (P), gene as ORF (G), and terminator (T). M: selection marker. InsUP and InsDOWN: genome integration targeting sequences upstream and downstream the cassette respectively. Letters A to M are assigned to each pre-designed 4nt overhang flanking the building blocks. Sequences of the overhangs are provided next to the respective position GGA: Golden Gate assembly. Figure adapted from Celinska et al., 2017

cassettes d'expression avec un ou deux TU seulement est également possible. Pour accroître l'utilité du système, une vaste bibliothèque de blocs interchangeables, composée de 64 briques, a été construite et caractérisée. La boîte à outils construite contient (i) neuf promoteurs, y compris des promoteurs natifs, hybrides et inductifs permettant une large gamme de force d'expression; (ii) cinq terminateurs, influençant différemment l'expression du gène; (iii) cinq marqueurs de sélection, y compris des marqueurs auxotrophes, des gènes de résistance aux antibiotiques et des gènes métaboliques; (iv) cinq séquences pour l'insertion au génome, trois d'entre eux ciblent des gènes impliqués dans le métabolisme lipidique de la cellule et les deux autres sont les éléments zeta classiques utilisés pour l'intégration aléatoire dans le génome de *Y. lipolytica* différant dans le site de reconnaissance enzymatique qu'ils présentent pour libérer la cassette du vecteur; et (v) trois gènes codant pour les protéines fluorescentes. Les différents éléments de la boîte à outils ont été testés à l'aide d'une protéine fluorescente rouge comme rapporteur de l'expression génétique (Larroude et al., 2019). En construisant des

souches capables de produire des caroténoïdes et des souches capables d'utiliser le xylose comme source de carbone, le système GG s'est avéré très efficace pour la construction rapide de souches avec des voies métaboliques complexes (Larroude et al., 2019, 2017). En outre, la capacité combinatoire de cette technique a été utilisée pour mélanger les promoteurs sur la voie à trois gènes pour la production hétérologue de caroténoïde. Cette technique a permis l'identification rapide de la meilleure combinaison gènes-promoteurs pour chaque position qui assure la meilleure capacité de production de la voie (Larroude et al., 2017). Les nombreux avantages de cette méthode GG augmentent fortement les possibilités d'ingénierie métabolique de *Y. lipolytica*. La bibliothèque réutilisable de fragments d'ADN caractérisés, ainsi que la capacité d'expansion illimitée de la bibliothèque et le brassage efficace des pièces permis par l'approche combinatoire de la méthode, permet un assemblage rapide d'une grande variété de briques biologiques. La capacité d'insérer trois gènes dans le génome en une seule cassette réduit le nombre de cycles de transformation nécessaires pour construire des voies complexes et réduit les marqueurs de sélection nécessaires. En outre, la cassette peut être introduite dans le génome par HR, permettant la suppression d'un gène en même temps, de sorte qu'une voie concurrente peut être réduite en même temps que la voie souhaitée est exprimée.

Pour l'édition du génome, un système CRISPR-Cas polyvalent et modulaire a été mis en place. Le vecteur contenant le système est construit par la méthode de GG, permettant ainsi le changement facile des marqueurs, promoteurs de Cas et gRNAs, et le recyclage des briques GG (Figure 44). Un ensemble de six vecteurs plateforme CRISPR-Cas9, chacun contenant de marqueurs de sélection différents, a été construit pour introduire facilement n'importe quel gRNA par assemblage GG. La boîte à outils développée peut être utilisée dans un large éventail de différents antécédents génétiques de *Y. lipolytica*, y compris les souches de type sauvage (Larroude et al., 2020). En outre, sur la base de la même conception vectorielle, un système CRISPR-Cpf1 a été construit et s'est avéré fonctionnel pour la modification du génome de *Y. lipolytica*.

L'assemblage modulaire rapide de voies synthétiques complexes et la capacité d'édition efficace du génome, et à un site spécifique, élargissent l'éventail des études possibles qui peuvent être menées dans *Y. lipolytica*. Les deux outils sont mises à la
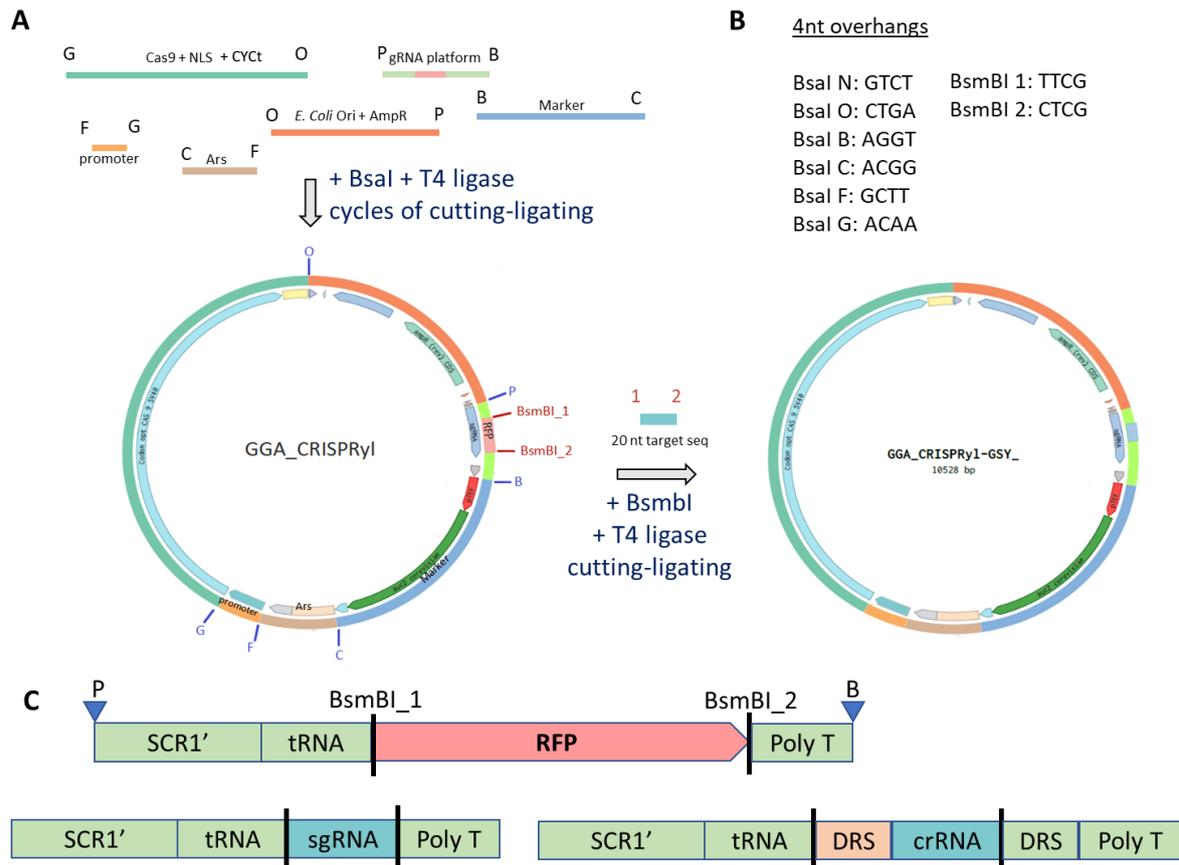
Figure 44. Modular CRISPR strategy based on GG assembly

A- Schematic representation of the modular two-step cloning strategy for the construction of CRISPR-Cas vector. In a first step, the building blocks are assembled to construct a platform vector containing a chromophore (RFP) at the place of the gRNA. In the second step, the RFP is released by BsmBI and the fragment with the 20nt specific to the gRNA are incorporated. Thus, only white clones are screen after cloning. **B-** The 4nt overhangs specific for each DNA fragment are listed. It is also specified the enzyme that is used to release the fragment from the donor vector. **C-** Representation of the structure of the DNA block harbouring the RFP (gRNA platform). It shows the places where BsmBI cuts, with the consequent release of the RFP. The sites P and B correspond to the 4nt overhangs produced after BsaI digestion and used to construct the platform vector. In the bottom part of the figure, the obtained structures after the cloning of the gRNA-specific sequence are represented. The left part corresponds to the Cas9 gRNA and the right to the Cpf1 gRNA. The gRNA and the endonuclease are the only differences between these two systems. RFP: Red Fluorescent Protein.

disposition de la communauté par Addgene et des protocoles détaillés ont été publiés dans les articles scientifiques (Larroude et al., 2020, 2019). Ces nouveaux outils déve-loppés ont permis la construction rapide de souches développées sur la deuxième partie de ce travail.

Avec le but de construire un châssis de *Y. lipolytica* avec un pool accru de AAAs, en particulier de Phe et Tyr, et d'élargir les connaissances sur la régulation et la fonc-tionnalité de la voie de synthèse des AAAs (Figure 45) dans cette levure, une grande série de souches a été construite.
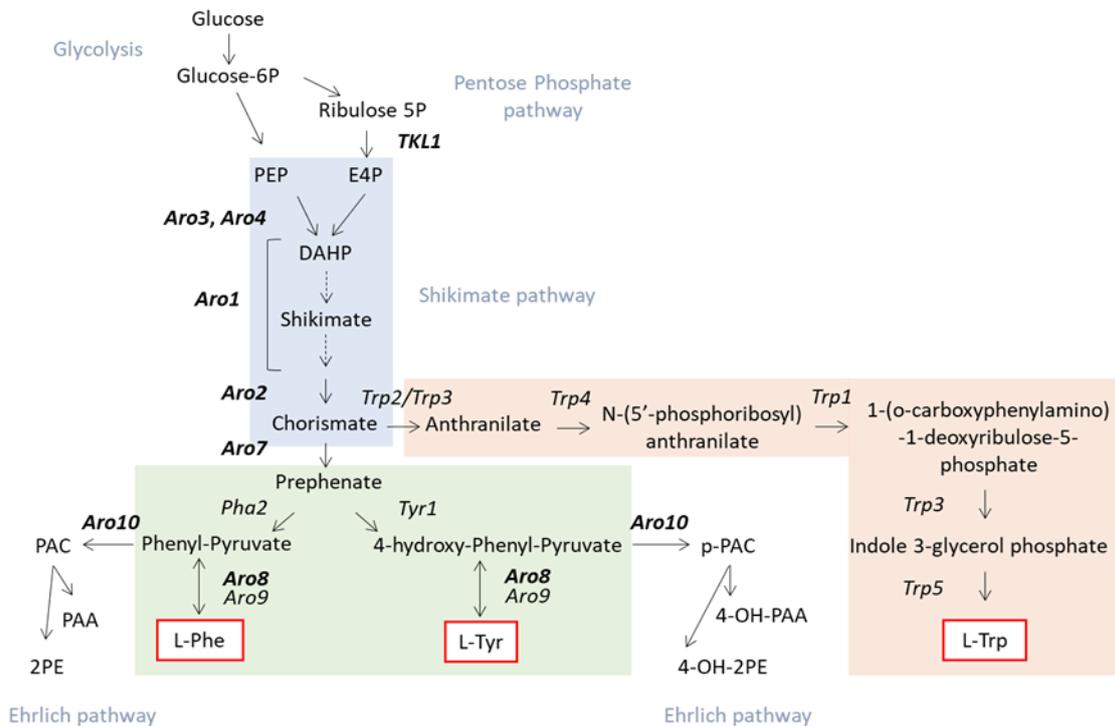
Figure 45. Aromatic amino acid pathway.

Representation of the AAA pathway divided into three parts: (1) the shikimic acid pathway (blue), (2) the L-Trp branch (red), and (3) the L-Tyr and L-Phe branches (green). The precursor availability through glycolysis and pentose phosphate pathway as well as the catabolism through the Ehrlich pathway are shown. Genes involved in the pathway are indicated next to the arrow representing the reaction. Genes shown on bold letters were involved in the construction of engineered strains presented during this work. Phe: phenylalanine; Tyr: Tyrosine; Trp: Tryptophan. 2PE: phenylethanol; PAA: phenylacetic acid; 4-OH-2PE: 2-(4-hydroxyphenyl)ethanol; 4-OH-PAA: 4-hydroxyphenylacetic acid

Afin de détecter les changements dans la voie des AAAs, trois approches ont été testées au cours de ces travaux : un biosenseur basé sur promoteur inductible, une molécule colorante dont le précurseur est le Trp et une méthode HPLC. La première, basée sur la région promotrice de *ARO10*, induite par Phe (Celińska et al., 2015b), n'a produit aucun signal détectable dans les conditions testées. De plus, aucune structure d'UAS susceptible d'être conçue pour augmenter la réponse du biosenseur n'a été mise en évidence dans la séquence du promoteur. La seconde approche, utilisant le PVA comme molécule verte rapportrice, a permis de démontrer les modifications sur le flux de la voie AAA à travers différentes intensités vertes dans la culture. Toutefois, le système n'était pas adapté à l'analyse à haut débit des souches et n'était pas pratique pour la détection des modifications du pool de Tyr et de Phe du fait que le PVA utilise le Trp comme précurseur. Ces deux premières méthodes ont mis en évidence le besoin d'un vecteur réplicatif stable pour *Y. lipolytica*, pour éviter l'insertion génomique de ces systèmes, et le vecteur CRISPR pourrait être testé pour une telle construction. Enfin, la

HPLC a été choisie comme méthode précise pour évaluer les modifications apportées au pool d'AAAs, bien qu'elle ne soit pas particulièrement adaptée à l'identification des souches à haut débit.

Pour la construction de souches modifiées, les gènes impliqués dans la voie des AAAs ont été adaptés pour GG et exprimés dans *Y. lipolytica*. Les résultats ont mis en évidence que la voie du shikimate doit être entièrement surexprimé pour avoir une augmentation significative du pool de Phe et Tyr (Figure 46). De plus, l'expression d'enzymes de *S. cerevisiae* insensibles au rétro-régulation (scAro4^{K229L}, scAro3^{K222L} et scAro7^{T226I}) augmente d'avantage la production de Phe et Tyr (Figure 47). Au contraire, la surexpression de *TKL1*, destinée à augmenter le précurseur E4P, n'a pas augmenté la production (Figure 47). Cela peut être dû au fait que cette enzyme peut avoir une préférence par la réaction opposée réduisant ainsi l'E4P disponible pour rentrer dans la voie du shikimate. Finalement, entre les souches construites, une produisait des quantités plus élevées de Phe et de Tyr que les autres (Figure 47). Cette souche, nommée JMY8175, exprime les gènes : *ARO1-ARO2-scARO4^{K229L}*-scARO3$K222L^{-scARO7}T226I$.
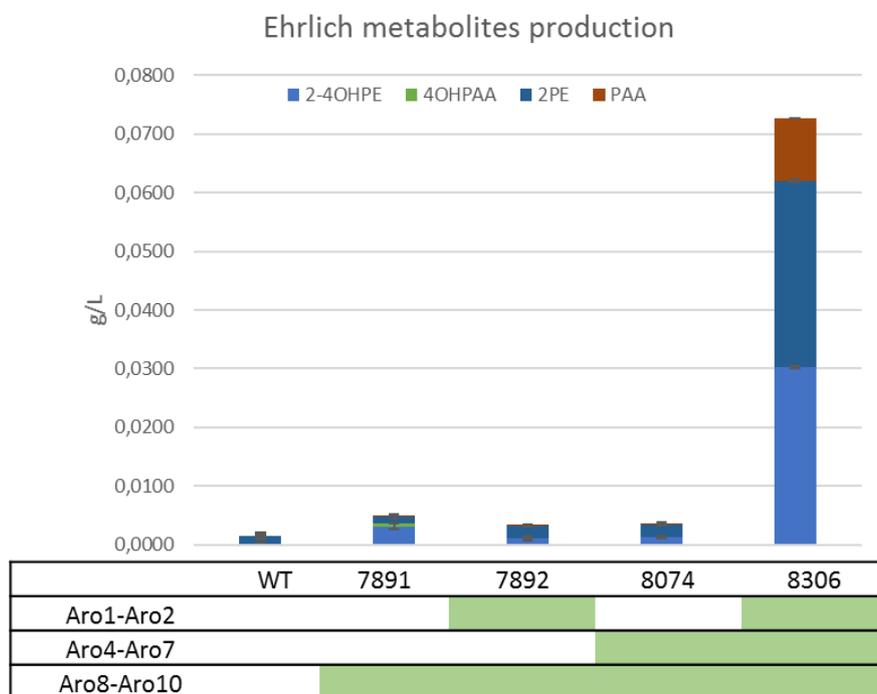

Figure 46. Evaluation of the AAA pathway of engineered strains.

Ehrlich metabolites production is used as reporter of the flux through the AAA pathway. The genotype of the constructed strains is indicated by the green-filled cases below their numbered-name. The numbers above the genotype indicate their collection number (JMYnumber). WT: wild-type strain. 2-4OHPE: 2-(4-Hydroxyphenyl)ethanol; 4OHPAA: 4-Hydroxyphenylacetic acid; 2PE: Phenylethanol; PAA: phenylacetic acid.

Cette première souche châssis d'AAA peut encore être améliorée, par l'ingénierie métabolique en ajoutant des gènes qui améliorent la quantité des précurseurs disponibles ou en éliminant des voies compétitrices, mais aussi par la modification des conditions de culture, comme par exemple les sources de carbone utilisées.
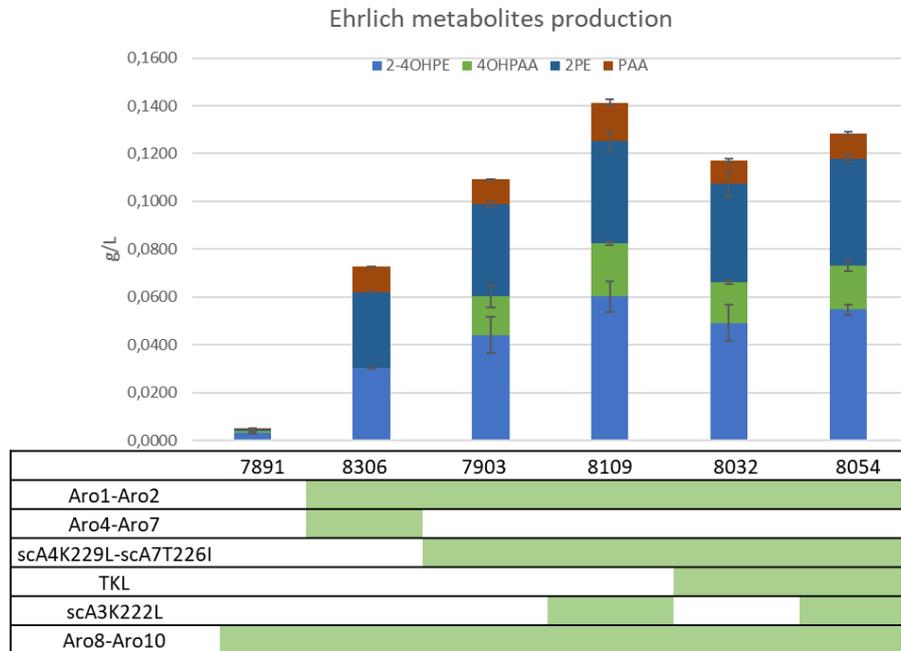

Figure 47. Evaluation of AAA the pathway on engineered strains overexpressing S. cerevisiae deregulated forms of Aro4 and Aro7 enzymes.

Ehrlich metabolites production is used as reporter of the flux through the AAA pathway. The genotype of the constructed strains is indicated by the green-filled cases below their numbered-name. The numbers above the genotype indicate their collection number (JMYnumber). 2-4OHPE: 2-(4-Hydroxyphenyl)ethanol; 4OHPAA: 4-Hydroxyphenylacetic acid; 2PE: Phenylethanol; PAA: phenylacetic acid.

Afin d'augmenter les connaissances sur la régulation de la voie des AAAs chez *Y. lipolytica*, la régulation des enzymes Aro4 et Aro7 par Phe et Tyr a été testée. Pour ce faire, un essai enzymatique a été effectué avec l'enzyme de type sauvage et, par la suite, des mutations ponctuelles ont été effectuées dans ces enzymes, à partir des mutations qui se sont révélées avoir un effet de dérégulation chez *S. cerevisiae* (Luttik et al., 2008). Les résultats ont montré que l'ARO4 de *Y. lipolytica* est réprimé par les composés finaux de la voie, Phe et Tyr, et que la mutation effectuée (K221L) évite cette régulation, ce qui entraîne une enzyme plus active. Cette forme dérégulée a été signalée très récemment (Palmer et al., 2020), et elle est également conforme à la régulation décrite sur l'enzyme *S. cerevisiae* (Luttik et al., 2008). Au contraire, Aro7 ne semble pas être régulé par Phe ou Tyr et la mutation effectuée n'a aucun effet sur l'activité de l'enzyme. Il s'agit d'un profil différent de l'activité de l'enzyme de *S. cerevisiae*. Au cours de cet

essai, on a également observé que la co-expression d'Aro4 et d'Aro7 avait un effet nocif sur la production de Phe et Tyr. Cela pourrait être dû à un effet auxotrophie de Tyr, néanmoins, ce phénomène doit être étudié davantage. Cet effet négatif a découragé l'utilisation de ces enzymes pour la construction de la souche châssis et, à la place, les formes dérégulées de *S. cerevisiae* ont été conservées. L'étude réalisée dans le cadre de cette thèse est, à ma connaissance, la première ayant comme but de mieux comprendre la régulation de la voie des AAAs chez *Y. lipolytica.*

Enfin, les souches de châssis AAA ont été utilisées pour la synthèse *de novo* de composés à valeur ajoutée endogènes et hétérologues dérivés de Phe et Tyr : 2-phény-léthanol, un composé aromatique au parfum de rose, naringenin, un flavonoïde, et mé-lanine, un protecteur UV et ROS (espèces réactives de l'oxygène). La production *de novo* de 2PE a été possible après l'ajout de deux gènes supplémentaires, *ARO8* et *ARO10*, à la souche châssis, atteignant 0,18 g/L. Il a été démontré que le glycérol comme source de carbone était plus efficace que le glucose, peut-être en raison de la capacité du glycérol d'augmenter l'E4P. Même si la quantité produite par *Y. lipolytica* est toujours inférieure à la production par *S. cerevisiae,* la production de 2PE peut être augmentée en modifiant l'équilibre entre les métabolites Ehrlich (Figure 48), en plus d'autres améliorations dans la souche châssis.

En ce qui concerne la naringenin et le resvératrol, la production obtenue a été très faible en raison d'une faible expression des gènes hétérologues et de la perte de certains gènes de la voie des AAAs. Cependant, comme les causes ont été identifiées, les cor-rectifs rapides sont possibles. En outre, le fait que même avec une faible augmentation du pool des AAA et une faible expression des gènes la naringenin est produit, est en-courageant pour poursuivre avec d'autres modifications de la voie afin d'augmenter la production. Ainsi, en plus d'améliorer la souche châssis, une mélange aléatoire des gènes orthologues et de promoteurs par GG est envisagé.

La souche modifiée s'est révélée être une bonne productrice de pyomélanine. Le pigment brun a été identifié et caractérisé par une analyse physicochimique et par la surexpression et la délétion d'un gène impliqué dans sa voie biosynthétique. Une souche d'hyperproductrice a été identifiée, capable de produire de la pyomélanine beau-coup plus rapidement que les autres souches. Les conditions de culture et les méthodes
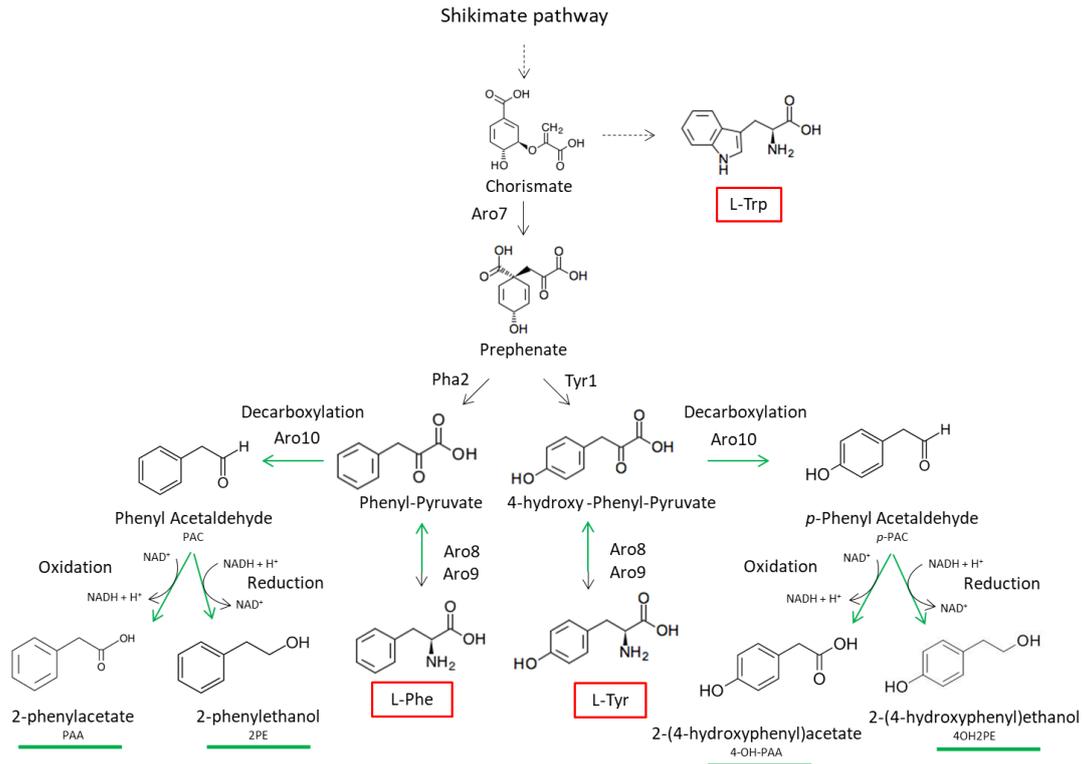
Figure 48. Ehrlich pathway

Representation of the Ehrlich pathway from Phe and Tyr (green arrows). In green lines are highlighted the four Ehrlich metabolites obtained. In red squares are shown the AAAs.

d'extraction peuvent être optimisées pour favoriser la production de ce pigment, néanmoins, 1 g/L a été obtenu pendant ce travail.

Même si des travaux très récents ont apporté des modifications mineures à la voie des AAAs de *Y. lipolytica* (Lv et coll., 2019b, 2019a; Palmer et al., 2020), aucun d'eux n'a visé la construction d'une souche châssis avec un pool accru de AAAs, ni l'exploration de la réglementation et l'ingénierie de cette voie.

Ce travail a élargi avec succès la boîte à outils dédiée à *Y. lipolytica* avec deux techniques robustes, rapides et polyvalentes, l'une pour l'assemblage de l'ADN et l'autre pour l'édition du génome à un endroit spécifique. Ces techniques se sont révélées très utiles pour la construction rapide d'usines cellulaires optimisées de cette levure. L'optimisation de la voie des AAAs, peu explorée chez *Y. lipolytica*, élargit la portée des applications possibles de ce microorganisme pertinent sur le plan industriel et en même temps augmente les connaissances sur la façon dont cette voie métabolique peut fonctionner dans cette levure. Les informations acquises vont alimenter le cycle itératif nommée Design-Build-Test-Learn (designer-construire-tester-apprendre) (Figure 49),

et vont permettre de construire une prochaine génération d'usines de cellules aromatiques plus performantes ainsi que de concevoir d'expériences améliores pour la détection des produits et de culture de souches.
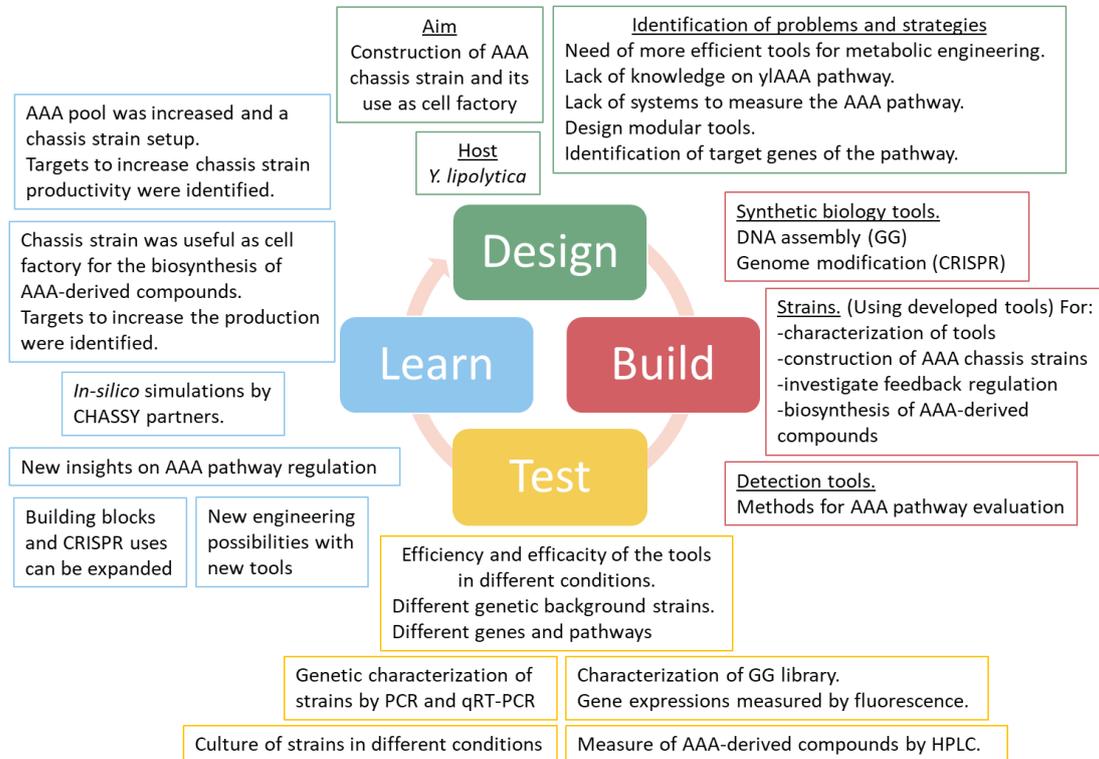

Figure 49. The DBTL cycle during this work

The different strategies of the presented work are part of a DBTL cycle. The acquired knowledge will feed the next round of the cycle with the aim of improving the efficiency of the cell factories and the basic knowledge on *Y. lipolytica* physiology. The DBTL cycle during this work

Titre : Développement d'outils de biologie synthétique et de souches châssis de *Yarrowia lipolytica* pour la production de molécules aromatiques

Mots clés : Biologie synthétique, ingénierie métabolique, *Yarrowia lipolytica*, composés aromatiques, CRISPR, GoldenGate

Résumé : La production rapide et à moindre coûts de molécules par voie microbiologique nécessite le développement de souches châssis robustes, polyvalentes et bien caractérisées. De plus, la disponibilité d'outils de biologie synthétique performants et normalisés est cruciale pour construire ces souches de manière efficace. Ces organismes rationnellement conçus pour avoir une disponibilité optimisée de précurseurs, doivent servir de base pour la synthèse d'une large gamme de molécules. Lors de ce projet, des outils de biologie synthétique dédiés à la levure *Yarrowia lipolytica* ont été développés. Une technique de clonage modulaire basée sur la méthode Golden Gate (GG) a été mise en place pour assembler jusqu'à 13 blocs via une réaction en une seule étape. Le kit développé, composé de 64 briques, s'est révélé être un outil puissant pour l'ingénierie complexe de souches et est parfaitement adapté pour des approches combinatoires. De plus, un système polyvalent et modulaire de vecteur CRISPR-Cas9, basé sur la méthode GG, a été mis en place pour l'édition génomique. Six vecteurs plateforme ont ainsi été construits pour introduire facilement les gRNA et pour être utilisé dans n'importe quel fond génétique, y compris les souches sauvages. Ces deux boîtes à outils sont mises à disposition de la communauté par Addgene. Ces nouveaux outils ont été utilisés pour la construction de souches châssis de *Y. lipolytica* optimisées pour la synthèse d'acides aminés aromatiques(AAA), précurseurs de nombreuses molécules aromatiques. Cette souche plateforme a été transformée en usine cellulaire pour la production de composés aromatiques d'intérêt industriel.
Ainsi, grâce à l'expansion des outils d'ingénierie pour *Y. lipolytica* et à la construction d'un châssis optimisé pour la production des AAA, ce travail élargit considérablement la portée des études et des applications possibles pour ce microorganisme d'intérêt industriel.

Title : Development of synthetic biology tools and construction of *Yarrowia lipolytica* chassis strains for aromatic molecules production

Keywords : Synthetic biology, metabolic engineering, *Yarrowia lipolytica*, aromatics compounds, CRISPR, GoldenGate

Abstract : To produce biobased products by microbial cell factories in a time and cost-effective way, the development of robust, versatile and well-characterized chassis strains is essential. The availability of efficient, standardized and characterized synthetic biology tools is crucial to build these chassis in a fast and efficient manner. These rationally engineered strains, based on an optimized supply of precursors, provide the foundations for the synthesis of a wide range of products.
During this project, synthetic biology tools dedicated to the yeast *Yarrowia lipolytica* were developed. A modular cloning technique based on Golden Gate (GG) was set-up for assembling up to 13 modular building blocks via a single-step, one-pot reaction. The toolkit, composed of 64 bricks, proved to be a powerful tool for rapidly engineer *Y. lipolytica* strains with complex pathways and is perfectly suited for combinatorial approaches. In addition, based on the GG assembly method, a versatile and modular CRISPR-Cas9 system was set up for genome editing. A set of six platform vectors was constructed to easily introduce any gRNA and to be used in a wide range of different *Y. lipolytica* genetic backgrounds including wild-type strains. Both toolkits are available to the community through Addgene. These newly developed synthetic biology tools were used for the construction of *Y. lipolytica* tailor-made chassis strains optimized for the synthesis of aromatic amino acids (AAA), which are the precursors of many aromatic compounds. This platform strain was further engineered as a cell factory to produce aromatic compounds with industrial interest, 2-Phenylethanol, naringenin and melanin, as proof of concept.
Thus, through the expansion of *Y. lipolytica* toolbox and the construction of an AAA-chassis strain, this work greatly enlarges the scope of possible studies and applications for this industrially relevant microorganism.