Review

Yarrowia lipolytica strains and their biotechnological applications: how natural biodiversity and metabolic engineering could contribute to cell factories improvement

Catherine Madzak [1]*

[1] Université Paris-Saclay, INRAE [§], AgroParisTech, UMR SayFood, F-78850, Thiverval-Grignon, France

* Correspondence: Catherine.Madzak@inrae.fr

[§] INRAE is France's new National Research Institute for Agriculture, Food and Environment, created on 01/01/2020 by the merger of INRA, the National Institute for Agricultural Research, and IRSTEA, the National Research Institute of Science and Technology for the Environment and Agriculture.

Abstract: Among non-conventional yeasts of industrial interest, the dimorphic oleaginous yeast *Yarrowia lipolytica* appears as one of the most attractive for a large range of white biotechnology applications, from heterologous proteins secretion to cell factories process development. The past, present and potential applications of wild type, traditionally improved or genetically modified *Yarrowia lipolytica* strains will be resumed, together with the wide array of molecular tools now available to genetically engineer and metabolically remodel this yeast. The present review will also provide a detailed description of *Yarrowia lipolytica* strains and highlight the natural biodiversity of this yeast, a subject little touched upon in most previous reviews. This work intends to fill this gap by retracing the genealogy of the main *Yarrowia lipolytica* strains of industrial interest, by illustrating the search for new genetic backgrounds and by providing data about the main publicly available strains in yeast collections worldwide. At last, it will focus on exemplifying how advances in engineering tools can leverage a better biotechnological exploitation of the natural biodiversity of *Yarrowia lipolytica* and of other yeasts from the *Yarrowia* clade.

Keywords: white biotechnology; metabolic engineering; non-conventional yeast; oleaginous yeast; cell factory; heterologous expression; biodiversity; *Yarrowia lipolytica*; *Yarrowia* clade; GMO

1. Introduction

A major challenge for our societies is to replace polluting technologies, based on fossil fuels, with clean ones, based on renewable resources. White biotechnology, using microorganisms and their enzymes to manufacture compounds of industrial interest (chemicals, biomaterials, biofuels, pharmaceuticals, feed, food), has an important role to play in this transition. This rapidly developing field aims to design industrial processes more environmentally friendly and making use of agricultural, forest and industrial waste or byproducts. Among the microorganisms amenable for such industrial applications, yeasts cells present the cumulated advantages of high growth capacity, easy genetic manipulation and presence of a eukaryotic organisation allowing posttranslational processing, vesicular secretion and subcellular compartmentalization.

After having drawn some industrialists' attention as early as the 1950s, the non-conventional oleaginous yeast *Yarrowia lipolytica* has been recognized since several decades, as a powerful host for heterologous protein expression, secretion and surface display. The development of sequencing and genetic engineering tools, combined with an increasing knowledge of its metabolism, have then facilitated the complex engineering of the metabolic pathways of this yeast for various applications. Since nearly two decades, numerous laboratories throughout the world have chosen *Y. lipolytica* as a chassis for designing microbial cell factories. White biotechnology applications of this yeast include notably single

cell oil production, whole cell bioconversion and upgrading of industrial wastes. This history of use will be briefly resumed, but the present publication intend to put a new emphasis on the choice of available *Yarrowia lipolytica* strains and the natural biodiversity of this species. This review will present the various *Y. lipolytica* strains of industrial interest, retrace their genealogy and resume their preferred applications. As the numerous molecular tools available for the genetic engineering of this yeast have already been abundantly described elsewhere, this review will focus on those that leverage a better biotechnological exploitation of the natural biodiversity of this species and, possibly, of other yeasts from the *Yarrowia* clade.

2. Everything you always wanted to know about *Yarrowia lipolytica* (briefly resumed)

2.1. Natural habitats and safety

   Y. lipolytica is a Crabtree-negative ascomycete yeast (class: Saccharomycetes, order: Saccharomycetales) that has been at first noticed for its remarkable lipolytic and proteolytic capacities. In accordance to these high levels of secreted enzymatic activities, wild-type isolates of this yeast generally originate from lipid-rich and/or protein-rich environments, notably from meat and dairy products (especially fermented ones, such as dry sausages and cheeses) and from sewage or oil-polluted waters [1, 2]. In the last decades, the range of ecosystems from which *Y. lipolytica* strains has been isolated has broadened to encompass very diverse habitats, from marine waters, salt marshes and soils (especially oil-polluted ones) to a variety of consumable products (including fruits, vegetables or seafood) and even the excreta of insects or vertebrates that consume them [1, 3-6]. This species thus appears to exhibit a rather ubiquitous distribution, in the natural world as well as in man-made extreme environments.

   Since only a decade, *Y. lipolytica* is also considered as belonging to the normal human mycobiota, being found notably in the mouth and respiratory tract of adults, especially of diabetic people [7]. This yeast is also sometimes seen as a possible opportunistic emerging pathogen, since its biofilm formation capacity can be responsible of rare cases of catheter-related candidaemia [1, 7]. Despite this, *Y. lipolytica* is classified as a Biosafety Level (BSL) 1 microorganism by the Public Health Service (USA). It is also recognized as a "microorganism with a documented use in food" by the International Dairy Federation (IDF) and the European Food and Feed Cultures Association (EFFCA), and as a "recommended biological agent for production purposes" by the European Food Safety Authority. This yeast has also gained a GRAS (generally recognized as safe) status, from the many GRAS notifications for its various applications that have been approved by the USA Food and Drug Administration (FDA) [1, 7].

2.2. Main characteristics

2.2.1. Physico-chemical conditions for growth

   Unlike most other hemiascomycetous yeasts, *Y. lipolytica* is an obligate aerobe, for which oxygen concentration constitute a limiting factor for growth. Its temperature limit is in the range of 32 to 34°C for most strains, although a very few can grow as high as 37°C. Most strains can be considered as psychrotrophic since they exhibit a residual growth when kept at 4-5°C. The preferred growth temperature is however in the range of 25 to 30°C [8]. *Y. lipolytica* is able to grow at a large range of pHs: most strains can be cultivated at pH 3.5 to 8.0 and a few can tolerate lower ones (2.0) or even very high pHs (9.7) [8]. Accordingly to its presence in salty environments and foods, *Y. lipolytica* tolerates high salt concentrations, such as 7.5% NaCl for most strains and as high as 15% NaCl for a few of them [8]. This yeast is also known to be able to adsorb metallic atoms and has therefore been proposed for bioremediation of wastes containing heavy metals such as Cr, Fe, Ni, Cu, Zn, and Cd [8].

2.2.2. Ploidy and morphology

This yeast is heterothallic, with two mating types Mat A and Mat B, and natural isolates are in most cases haploid [9, 10]. The mating frequency of two natural Mat-compatible strains is very low, but the resulting diploid state is stable under laboratory growth conditions [11]. Such hybrids exhibit a very low fertility, a problem that was alleviated through inbreeding programs to allow the establishment of the first genetic maps [9]. High sporulation rates can be obtained on peculiar media (yeast extract/malt extract or V-8 juice media, media with 1.5% sodium citrate as sole carbon source) and the shapes of asci and ascospores exhibit some strain-dependant variations [8]. Wild-type isolates of *Y. lipolytica* can present a large variety of colony aspects, ranging from smooth and glossy to strongly wrinkled and mat. This diversity reflects the fact that *Y. lipolytica* is a dimorphic yeast that can grow either as round multipolar budding cells, pseudohyphae (budding cells remaining attached) or mycelia with septate hyphae, depending on growth conditions [9, 10, 12]. This possibility of growth under different forms (dimorphic switch) is of practical importance regarding biotechnological applications of *Y. lipolytica*, since monitoring all environmental parameters will be crucial for the control of cellular morphology, from which the optimization of the bioprocess could depend on [13].

2.2.3. Carbon sources

Y. lipolytica is able to use as carbon source a large array of substrates of either hydrophilic or hydrophobic nature [9, 10]. Water-soluble carbon sources include only a few sugars (glucose, fructose, mannose) but also glycerol and, to a lesser extent, organic acids and alcohols. The long-prevailing belief that *Y. lipolytica* could use only some hexoses, but no pentose, as sole carbon source has however been recently undermined in experiments on xylose assimilation by some strains, as will be reported in Sections 2.3.3 and 3.2.1. Water-insoluble carbon sources comprise fatty acids, triglycerides and alkanes. Remarkably, the engineering of *Y. lipolytica* metabolism for the use of alternative substrates has been initiated very early in its history of genetic manipulation: heterologous expression of *Saccharomyces cerevisiae SUC2* gene was used more than three decades ago to confer the ability to grow on sucrose to some of the most used laboratory strains [14]. Genetic engineering of *Y. lipolytica* for use of agricultural and industrial waste as alternative substrates for white biotechnology applications has become an important and rapidly developing research field, as will be exemplified further in this review. This overall versatility that this yeast allows in the choice of possible substrates represents a valuable asset for the development of bioprocesses involving *Y. lipolytica*, especially those based on the valorisation of by-products or waste.

2.2.4. Secretion pathway

The two more prominent characteristics of *Y. lipolytica* are its very efficient secretion pathway and its outstanding lipid storage capacity. Consequently, this yeast has become a research model in the domains of protein secretion and lipid metabolism [15, 16]. The study of vesicular protein secretion in *Y. lipolytica* has demonstrated that the translocation of the nascent protein into the endoplasmic reticulum (ER) was mainly co-translational, as in the secretion pathway of mammals [15]. This peculiarity, contrasting to the situation in *S. cerevisiae* and most yeasts for which secretion is mainly post-translational, allows *Y. lipolytica* to be very efficient in the folding and secretion of large and/or complex heterologous proteins and has contributed to its success as a heterologous production host [17-19]. In addition, *Y. lipolytica* is one of the few yeasts, with *Pichia pastoris*, which lacks an ⊚-1,3-mannosyltransferase, a factor that limits the amount of excessive mannosylation of secreted heterologous glycoproteins and constitutes a valuable asset for the production of therapeutic proteins [20, 21].

2.2.5. Lipid storage

As an oleaginous yeast, *Y. lipolytica* can naturally accumulate lipids up to 30 to 50% of the cell dry weight (CDW), depending not only on each wild-type isolate genetic background but also on the carbon source used and the growth conditions. This lipid accumulation can reach up to 90% of CDW through genetic engineering, in obese *Y. lipolytica* cells [16] (cf. Section 3.1.4). This remarkable lipid storage capacity results from both an efficient uptake of hydrophobic substrates from the medium and an effective *de novo* synthesis pathway for triacylglycerols (TAG). The physiological response of *Y.lipolytica* cells to the presence of hydrophobic substrates (such as alkanes, fatty acids or oils) consists in the production of biosurfactants (notably liposan), in a hydrophobization of the cell membrane and in the formation of protrusions on the cell surface [15]. These protrusions correspond to the hydrophobic binding structures of an interfacial transport system, composed of several dozens of multimeric protein complexes, which facilitate the uptake of hydrophobic compounds from the environment [16, 22, 23]. The very efficient secretion of the extracellular lipase *LIP2* also contributes to the effective uptake of lipids by this yeast, through a reduction in molecular weight of the hydrophobic substrates. The lipase family has known an expansion in *Y. lipolytica*, as in most oleaginous yeasts, with a total of 16 lipase genes. The storage lipids of *Y. lipolytica* consist mostly of TAG and sterol esters, more than free fatty acids (FFA), and accumulate in a specialized subcellular compartment, the lipid body (LB). These lipids can comprise as high as 80% of unsaturated fatty acids, which present some valuable health benefits. Notably, *Y. lipolytica* is the oleaginous yeast with the highest known percentage of linoleic acid (LA), namely more than 50% [16]. The lipid metabolism of this yeast is of particular relevance for some major white biotechnology applications, such as the production of single-cell oil (SCO) and of biofuel, and has been the subject of numerous reviews [16, 22, 24].

2.2.6. Genomic organization

The first *Y. lipolytica* strain to be completely sequenced and fully assembled and annotated, E150, which will be described in detail in Section 3.1.2, constitutes the reference strain for genome structure studies. Its genome of 20.5 Mb comprises six chromosomes which sizes range from 2.6 to 4.9 Mb [25, 26]. This genome size is almost twice those of most other yeasts, including *S. cerevisiae* (12 Mb) [26]. Several other genomic characteristics make *Y. lipolytica* clearly stand out from the crowd of other hemiascomycetous yeasts. Notably, the G/C content, of 49% in average and near 53% in the genes, and the proportion of intron-containing genes, of 15%, are strikingly higher than for other yeasts (respectively 38%, 40% and 5% in *S. cerevisiae*) [26, 27]. In contrast, the number of genes, although on the strong side of the range for hemiascomycetous yeasts, is not as high as may be inferred from the large genome size. Namely, *Y. lipolytica* totalizes 6703 genes, more than the 5807 ones from *S. cerevisiae* but less than the 6906 ones from *Debaryomyces hansenii*, which both have genomes of around 12 Mb [26].

Nine other *Y. lipolytica* strains of different genetic background have now been sequenced (cf. Sections 3 and 4.5) and the already assembled genomes show some chromosomal rearrangements compared to the reference strain, despite a nearly constant genome size [28, 29]. This is consistent with the previous observation, in karyotypic analyses, of an important polymorphism in the length of various chromosomes between different *Y. lipolytica* strains. Such a high level of chromosomal rearrangements between strains could explain the poor fertility that was observed for the hybrids [30].

Among yeasts, *Y. lipolytica* presents atypical ribosomal DNA units, with several rRNA gene clusters scattered on different chromosomes (six clusters on four chromosomes in E150). In addition, the 5S RNA gene is not included in those rDNA unit, but present as separated copies scattered throughout the genome [30]. These characteristics, like the co-translational secretion pathway mentioned above, are closer to those of mammals than to those of other yeasts, confirming the eccentric phylogenetic position of *Y. lipolytica* based on the comparison of 18S and 26S rDNA sequences [30]. Some expression vectors for *Y. lipolytica* genetic engineering make use of rDNA sequences as targeting elements for integration into the genome (cf. Section 4).

The first *Y. lipolytica* retrotransposon identified, Ylt1, was detected in the E150 ge-nome; this element can only be found in a few wild-type isolates and in their derivatives, such as the genetically modified (GM) E150 strain (cf. Section 3.1.2). Ylt1 belongs to the Ty3/gypsy group and is bordered by unusually large (more than 700 bp) long terminal repeats (LTRs) termed zeta sequences, which can also be found as solo elements in the genome [31]. The numbers of Ylt1 and of solo zeta sequences present in a genome vary for each Ylt1-bearing strain but is of at least 35 copies for the retrotransposon and more than 30 copies for the solo LTRs [31]. A number of other retrotransposons have since been identified in other *Y. lipolytica* strains [28, 32, 33], but the presence of Ylt1 in a genome is relevant for some metabolic engineering strategies, since zeta sequences have been used as targeting elements in some expression vectors or cassettes, as will be explained hereaf-ter (cf. Section 4.1.2) [34, 35]. In contrast to *S. cerevisiae*, *Y. lipolytica* does not bear any re-trotransposon of the Ty1/Copia group, which are usually abundantly found in eukaryotic genomes. Interestingly, the presence in some strains of several retrotransposons and LTR-like sequences near to RNA polymerase III-transcribed genes, which number is almost twice that in most other yeasts, seems to indicate that these retro-elements may have played an active role in the evolution of *Y. lipolytica* [28].

2.3. An impressive curriculum vitae: short review of past, present and future Y. lipolytica uses

2.3.1. Industrial applications of wild type or traditionally improved strains

The high potential of *Y. lipolytica* for industrial applications has been exploited since more than seventies years, at first in the fields of biomass and valuable metabolites pro-duction, using proprietary wild-type isolates or traditionally improved strains (mutants, strains issued from hybridizations and crossings) [1, 36, 37]. In the 1950s, and until the oil crisis of the 1970s, the British Petroleum Company (BP, London, UK) applied the alkane-degrading capacities of this yeast to the production of single-cell protein (SCP) from crude oil. The product was commercialized under the name Toprina G, for livestock feeding, and prompted numerous studies on *Y. lipolytica* safety that led to the first GRAS (generally recognized as safe) notification for a process using this yeast. In the 1970s, Pfizer Inc. (New York City, USA) applied *Y. lipolytica* to industrial citric acid production. This technology was repurchased in the 1990s by Archer Daniels Midland Company (ADM, Chicago, IL, USA), who is still (to the best of our knowledge) producing some citric acid from *Y. lipo-lytica*, using mostly corn or rape seed oil as substrates [1, 36, 37]. Other nowadays appli-cations of proprietary wild-type or traditionally improved *Y. lipolytica* strains include, in the field of food industry, erythritol production by Baolingbao Biology Co. (Yucheng, Shandong, China) and use of *Y. lipolytica* biomass as fodder yeast for farm and pet animals by Skotan SA (Chorzów, Poland), who also develop prebiotic/probiotic applications [1, 38].

The outstanding capacity of *Y. lipolytica* for degrading hydrocarbons, and especially alkanes, explains that wild-types isolates were frequently found in oil-polluted environ-ments and justifies the use of this yeast in bioremediation projects [36]. Although most of such researches are only at the laboratory level, two field studies were reported in Poland: the wild-type *Y. lipolytica* A-101 strain (cf. Section 3.2.2) has been successfully applied to *in situ* depollution of petroleum-contaminated soils at a fuel base [39] and of creosote-contaminated soils at a timber processing plant [40]. In the same field of bioremediation, a *Y. lipolytica* based starter for depolluting wastewaters is commercialized by Artechno (Isnes, Belgium). This product corresponds to freeze-dried *Y. lipolytica* cells and secreted lipase from selected highly lipolytic mutants of ATCC 48436 strain (cf. Section 3.2.1), ob-tained using chemical mutagenesis at INRA [41].

2.3.2. Commercial applications of genetically modified strains

In the 1980s, the newly developed technics of molecular biology rejuvenated the in-terest in *Y. lipolytica*, this time as an expression host for producing heterologous proteins [17]. Metabolic engineering of this yeast ensued rapidly, following the development of

transformation methods, shuttle vectors and non-leaky non-reverting auxotrophic strains [14]. As *Y. lipolytica* started, in the 2000s, to be recognized as a valuable host for recombinant protein production [18, 19], the YLEX kit for expression/secretion of heterologous proteins in this yeast was commercialized in 2006 by Yeastern Biotech Co. (Taipei, Taiwan - http://www.yeastern.com/). This kit includes vectors for expression/secretion and the Po1g strain, a GM derivative of W29 wild-type isolate, obtained at INRA (cf. Section 3.1.3). During the same period of time, *Y. lipolytica* other W29 derivatives have been established as commercial protein production platforms by Protéus (Sequens Group, Ecully, France) and Oxyrane UK (Manchester, UK), both also making use of INRA technologies [42].

With the continuous progress of genetic engineering technics, increasingly complex modifications of *Y. lipolytica* metabolism, like the introduction of complete heterologous metabolic pathways, could be performed. Proofs of concept of the use of this yeast as cell factory for the production of valuable compounds or as arming yeast for bioconversion processes are abundantly reported in the scientific literature since a few decades [42-44]. However, most of the proposed applications for these GM *Y. lipolytica* strains remain, until now, only at an exploratory stage and are not developed further to the industrial stage. This matter of fact could be attributed at least in part to social acceptance issues concerning GM microorganisms, especially in the domain of food applications. Until now, only a few commercial or industrial applications of GM *Y. lipolytica* strains can be reported [1, 37]. GM *Y. lipolytica* cell factories are presently used for industrial production of two kinds of food/feed additives: carotenoids and polyunsaturated fatty acids (PUFAs)-rich SCOs. The technology of carotenoids production by *Y. lipolytica* was developed by Microbia (USA) and then acquired by DSM (Heerlen, Netherlands) [37]. The technology of PUFAs-rich SCOs production by a heavily engineered *Y. lipolytica* strain derived from the ATCC 20362 wild-type isolate (cf. Section 3.2.1) was developed by DuPont (Wilmington, DE, USA) and was more particularly applied to industrial production of ω-3 eicosapentaenoic acid (EPA)-rich products [45, 46]. The use of GM *Y. lipolytica* for producing EPA-rich SCO was patented by DuPont (US2009/0093543A1) and the corresponding engineering work was described in details a few years later [45]. Remarkably, the development of this first commercially viable technology platform using GM *Y. lipolytica* was achieved using only the classical genetic engineering tools available at the time [46]. Two commercial products were derived from this technology platform: EPA-rich SCO for food and EPA-rich *Y. lipolytica* biomass for feed applications. The EPA-rich SCO has been marketed, as dietary supplement for human consumption, under the name New Harvest™, but only briefly (from 2010 to 2013) before the product was discontinued. New Harvest™ was advertised as the first vegetarian alternative to fish-based ω-3-rich oils, with no mention of its GM nature. No confirmation of social acceptance issues concerning the use of a GM microorganism was given by the company, concerning the product withdrawal. Since 2010, the EPA-rich *Y. lipolytica* biomass is marketed, in joint venture with AquaChile (Puerto Montt, Chile), as an ω-3 feed supplement for "harmoniously raised" salmon Verlasso™. DuPont research teams have also recently engineered the same ATCC 20362 strain for producing 2'-fucosyllactose (2'FL), the major human milk oligosaccharide (HMO), with the aim of developing a commercially viable cell factory platform for 2'FL and perhaps for other HMOs for infant formulas [47].

Another domain of successful applications for GM *Y. lipolytica* strains is the therapeutic use of recombinant enzymes: several enzyme replacement therapies (ERTs) based on this yeast are now marketed or on the edge to marketing stage [42, 43]. The first of these ERTs was developed by Mayoly Spindler (Chatou, France), in partnership with INRA and AzurRx Biopharma, Inc. (Brooklyn, NY, USA / Langlade, France), from a GM *Y. lipolytica* strain overexpressing *LIP2* [48]: this recombinant extracellular lipase is currently applied for the treatment of exocrine pancreatic insufficiency, and under Phase 2 clinical trial for two other fat malabsorption diseases (cystic fibrosis and chronic pancreatitis). More recently, Oxyrane (Ghent, Belgium), starting from Po1d strain (a GM derivative of W29 wild-type isolate, obtained at INRA - cf. Section 3.1.3), has established a proprietary *Y. lipolytica* engineering platform able to produce recombinant glycoproteins,

with the possibility of added mannose-6-phosphate (M6P) glycan residues [49], for treatment of different lysosomal storage diseases. The presence of M6P on therapeutic glycoproteins improves their internalization into the patient's cells and addresses them to lysosomes, their targeted subcellular site of action. A recombinant human acid α-glucosidase produced in *Y. lipolytica*, OXY2810, is currently marketed for use as ERT in Pompe disease, in which glycogen accumulates in the patient's tissues. A recombinant glucocerebrosidase (GCase) is in preclinical testing for treatment of Parkinson's disease and a GCase adorned with M6P glycans is in development for treatment of neuronopathic Gaucher disease, due to glucocerebroside accumulation in neuronal cells [42, 43]. Oxyrane also envision to apply their *Y. lipolytica* platform to developing new ERTs targeting other metabolic diseases.

2.3.3. Towards a bio-based economy: rewiring strain metabolism for alternative substrates

As said above, numerous proofs of concept of the use of GM *Y. lipolytica* for various potential white biotechnology applications can be found in the scientific literature. Their number is increasing rapidly: a search about "Yarrowia AND engineering" on PubMed website generates 17 results for the year 2010, an annual average of 30 results for the four following years, then an annual average of 64 results for the three following ones and finally an annual average of 105 results for the three last years, until 2020. For 2021, the search already generates 47 results already for the first trimester, which could be extrapolated to nearly 200 results if the rhythm of publication is maintained. These innovations have been abundantly described in many recent reviews [37, 38, 42-44, 50-53] and will only be resumed here and schematically depicted in Figure 1 that represents a state of the art for substrates availability and biotechnological applications for GM *Y. lipolytica* strains.
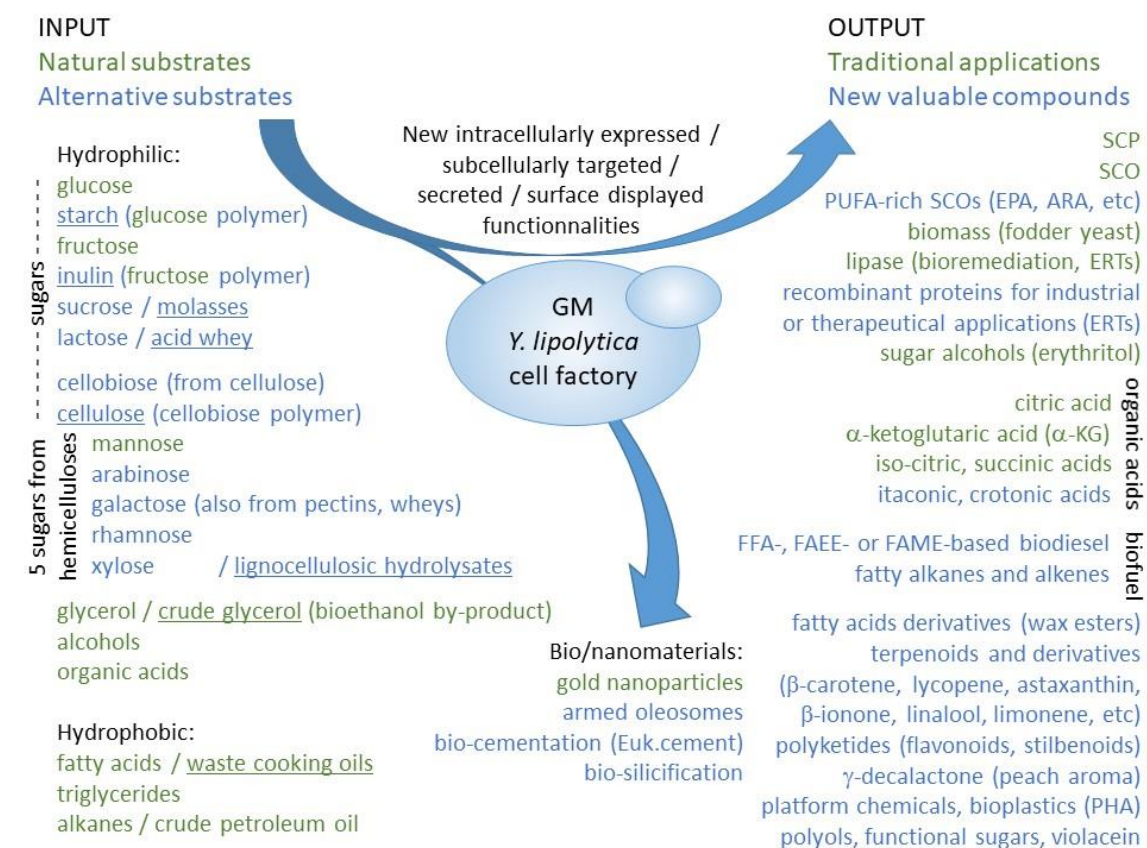

Figure 1. Schematic representation of genetically engineered *Y. lipolytica* cell factories as a generic "black box", with exploitable substrates as input and potential applications and products as output

Natural substrates and traditional applications of wild-type *Y. lipolytica* strains are indicated in green. Alternative substrates and new applications, requiring metabolic remodelling of *Y. lipolytica*, are indicated in blue (pentoses are indicated in blue, according to the fact that most *Y. lipolytica*

strains cannot use them as sole carbon source: cf. more details in Sections 2.3.3 and 3.2.1). Substrates issued from agricultural, forest and industrial waste or by-products, corresponding to an environmentally friendly concept of circular bio-economy, are underlined. Data compiled from several recent reviews [37, 38, 42-44, 50-53]. Abbreviations used, per order of occurrence in the figure: GM, genetically modified; SCP, single cell protein; SCO, single cell oil; PUFA, poly-unsaturated fatty acids; EPA, eicosapentaenoic acid; ARA, arachidonic acid; ERTs, enzyme replacement therapies; α-KG, α-ketoglutarate; FFA, free fatty acids; FAEE, fatty acid ethyl esters; FAME, fatty acid methyl esters; PHA, polyhydroxyalkanoates.

The metabolic remodelling of *Y. lipolytica* for the use of alternative renewable substrates as cheap carbon sources is aimed at valorising waste or by-products from various human activities or industries, in a sustainable development and circular bio-economy approach. The possibility to use raw materials as feedstock is also of major importance for reducing the production cost, in order to allow industrialization of processes that are still at the laboratory-scale. Besides the use of crude glycerol (a by-product of industrial production of bioethanol or soap) and of waste cooking oils (recycled from fast-food and other catering establishments), which both are natural substrates for *Y. lipolytica*, genetic engineering can also allow the use of lignocellulosic hydrolysates from agriculture, forestry and paper industry, of starch- or inulin-rich agricultural wastes, of molasses from sugar industry and of acid whey from cheese and yogurt industries [50, 51]. Interestingly, the Wroclaw University of Environmental and Life Sciences (UPWr, Poland) collaborated with INRA for the design of GM *Y. lipolytica* strains combining an obese phenotype with a wide substrate range [54]. Notably, these teams engineered GM derivatives of either W29 or A-101 wild-type isolates to produce at first obese strains with increased lipid biosynthesis and storage, and then to add several biomass-derived sugars (galactose, fructose, sucrose and inulin) to the range of substrates that these strains can effectively use (cf. more details on the resulting strains, YLZ150 in Section 3.1.4 and Y4779 in Section 3.2.2). The W29-derived YLZ150 strain, which can perform very efficient lipid biosynthesis from a wide range of biomass-derived sugars, is a valuable chassis for SCO or biofuel production from non-lipid renewable resources. In contrast, the A-101-derived Y4779 strain would be more adapted to developing sustainable processes for citric acid or polyhydroxy alcohol production [54].

However, a major challenge for white biotechnology is to select or engineer microorganisms of interest that can efficiently metabolize xylose, the main component from lignocellulosic material. Indeed, lignocellulosic hydrolysates, which constitute the preferred substrate in a circular bioeconomy, are a raw mixture of nutrients, notably pentoses and hexoses (mainly xylose and glucose), originating from the degradation of cellulose and hemicellulose from plant biomass. These hydrolysates can also contain significant amounts of some compounds, like organic acids and furfurals, which are toxic for most organisms but are rather well tolerated by *Y. lipolytica* [55]. This yeast has been regarded for a long time as unable to metabolize xylose, since the usual wild-type isolates were not able to use this pentose as sole carbon source. This drawback has prompted different research teams to engineer *Y. lipolytica* strains for xylose use, with the aim to develop sustainable SCO production platforms, as reviewed previously [55]. However, xylose consumption has rather recently been found to be strain-dependent (cf. Section 3.2.1) and omics studies have revealed a dormant pentose pathway in *Y. lipolytica* [56]. A recent collaborative work performed mainly at INRA has demonstrated that the overexpression of three genes from this dormant pentose pathway, in a Po1d genetic background, conferred to the recombinant strain the ability to grow on xylose as well as the parent strain on glucose, and was even more efficient than the overexpression of corresponding heterologous genes [55]. The same overexpression of native genes, performed in an obese derivative of Po1d strain, generated a strain able to accumulate lipids up to 67% of CDW, using xylose as sole carbon source. These two strains were also demonstrated to be able to grow efficiently in lignocellulosic hydrolysates, and could potentially be applied to sustainable production of xylitol and citric acid, or of SCO/biofuel, respectively [55]. In order to

achieve the same goals of developing sustainable industrial processes, the University of Texas, Austin (UT Austin, USA) has recently constructed GM strains able to use xylose as sole carbon source, by addition of a heterologous oxidoreductase pathway, followed by a starvation adaptation step, into a Po1f strain and one of its obese derivative, the E26 strain (cf. Section 3.1.4). The resulting XUS (xylose utilizing strains) were submitted to genome resequencing which revealed notably a duplication of the newly introduced *XYL1* and *XYL2* genes from *Scheffersomyces stipitis* to be a major factor for improved xylose consumption [57]. The PO1f XUS and E26 XUS strains were subsequently used in a strain mating strategy that allowed to combine their xylose consumption property with the capacity to synthetize various valuable compounds from several GM strains, as will be detailed in Section 4.4.2 [58].

2.3.4. Potential applications of genetically modified strains: bioproducts and biofuels

Besides the traditional biotechnological applications of *Y. lipolytica*, most of which have already been upgraded to industrial level (cf. Section 2.3.1), genetic engineering now allows to envision the production of a large range of valuable compounds. In addition, the production of natural metabolites or products of interest can also benefit of strain improvement obtained through metabolic engineering, either for increased yield or for use of alternative substrates, as has been exemplified for α-ketoglutaric acid (KGA) [59], for γ-decalactone (peach aroma) [60], for SCP [61] and for SCO [62], which composition can also be enriched in health-promoting ω-3 or ω-6 rich PUFAs, as seen above [45, 46] (cf. Section 2.3.2). The production of recombinant proteins for industrial or therapeutical applications represents an important research field that has been extensively reviewed previously [17-21]. The particular need to obtain more human-compatible recombinant glycoproteins for possible use as therapeutic agents has prompted the development of glyco-engineered (aka humanized) *Y. lipolytica* strains, which will be evoked in Section 3.1.6. In the domain of pharma/food industry, new metabolites of interest that can be produced using GM *Y. lipolytica* cell factories comprise notably a large range of terpenoids (including various carotenoids) and their derivatives [63-65], polyketides (including flavonoids) [64], riboflavin [58] and human milk oligosaccharides [47] (cf. Section 2.3.2). More recently, *Y. lipolytica* has also been applied to the production of a bacterial compound, violacein, a bis-indol purple pigment derived from the tryptophan pathway, which has numerous therapeutic, notably anticancer, properties [66]. GM *Y. lipolytica* are also considered as emerging cell factories for the production of several organic acids non-naturally produced by this yeast, of sugar alcohols (polyols) and of a number of functional sugars (including the naturally produced erythritol) which have a wide range of applications in different economic sectors [67, 68]. In the chemical industry domain, GM *Y. lipolytica* are an interesting source of several value-added bioproducts: wax esters used as advanced lubricants [69]; polyhydroxyalkanoates and polylactic acid homopolymer used as bioplastics [70, 71]; platform chemicals, like itaconic and crotonic acids (organic acids) [66, 72], ricinoleic acid (an unusual fatty acid) [73], triacetic acid lactone [58] or multi-purpose long chain dicarboxylic acids used notably in the synthesis of polyamides and polyesters copolymers [74].

Last but not least, GM *Y. lipolytica* cellular factories can constitute a particularly interesting source of biofuels, due to their high oleaginicity, to the possible use of inexpensive renewable carbon sources and to their robustness and good performance under stress, as was abundantly reviewed previously [75-78]. Increasing the oleaginicity of *Y. lipolytica* in order to obtain GM obese strains that can be used as chassis for producing various valuable lipid-related products and biofuels is an important research domain that will be reviewed in Section 3.1.4. Most researches on *Y. lipolytica*-produced biofuel focus on rewiring its metabolism for an increased accumulation of fatty acids and derivatives, since FFAs, fatty acid ethyl esters (FAEEs) and fatty acid methyl esters (FAMEs) can be used directly as biodiesel. Alternatively, another strategy is aimed at the production of hydrocarbon-based biofuels. Notably, *Y. lipolytica* was engineered to obtain the biosynthesis of pentane through heterologous expression of a soybean lipoxygenase [79] and of

larger odd-chain alka(e)nes through heterologous expression of the fatty acid photodecar-boxylase from *Chlorella variabilis* [80, 81]. Such studies demonstrate that GM *Y. lipolytica* cellular factories can also be used to produce renewable alkanes and alkenes.

Both of these engineering strategies, for fatty acid-based and hydrocarbon-based bio-fuel production, can benefit of recent methods of subcellular targeting of new functional-ities to different cell organelles. A research team from the Massachusetts Institute of Tech-nology (MIT, USA) engineered *Y. lipolytica* into a yeast biorefinery platform for sustaina-ble production of fuel-like compounds and oleochemicals, in which subcellular targeting of heterologous enzymes to the cytoplasm, the peroxisome and the ER allowed to generate FAMEs and fatty alkanes with tailored chain length [82]. Similarly, a team from the Huazhong University of Science and Technology (Wuhan, China) has targeted lipase de-pendent pathways to different lipid-related subcellular compartments (LB, peroxisome and ER), which allowed a substantial increase of FAMEs, fatty alkanes or fatty alkenes titers, compared to their cytosol-targeted engineered counterparts [83]. The same research team has also used an innovative method for the *in vivo* self-assembly of multienzyme complexes (MECs) in order to improve fatty acid-derived hydrocarbon production in *Y. lipolytica* cells. Their strategy was based on the simultaneous surface display of scaffoldin, a synthetic multiple cohesins backbone, and extracellular secretion of several dockerin-fused heterologous enzymes implied in the new metabolic pathway [84]. The highly spe-cific interaction of cohesin and dockerin domains allowed the spontaneous assembly of MECs carrying up to three enzymatic functions, in a combinational way. Namely, differ-ent proportional and positional effects could be genetically encoded in a customized scaf-foldin backbone by varying the copy number and orientation of the cohesin domains used. This methodology allowed an optimization of the co-immobilized enzymes efficiency, through a substrate channelling effect that generated a 17-fold enhancement in their initial reaction rate. The resulting MECs exhibited much higher conversion yields (71-84%) in alka(e)ne production than an equivalent cocktail of free enzymes (8-32%) [84].

2.3.5. Potential applications of genetically modified strains: nanoparticles and biomateri-als

A more unexpected domain of applications of GM *Y. lipolytica* cells is the production of biomaterials or nanostructures [42]. Previously, some wild-type *Y. lipolytica* strains had been applied to the biosynthesis of metallic nanoparticles for biomedical applications, no-tably NCIM 3589 strain for gold nanoparticles (AuNPs) production [85]. As the formation of AuNPs was directed by the melanin/pyomelanin produced by *Y. lipolytica* cells, a more recent strategy consists in engineering *Y. lipolytica* strains for enhanced production of these pigments that could then be purified and applied to AuNP synthesis. Notably, the synthesis of AuNPs mediated by pyomelanin purified from a GM W29 strain was recently reported, and a multifactorial statistical analysis of the process parameters allowed to fine-tune the size of these nanoparticles for different medical applications, such as imaging or drug delivery [86].

Prior to its use for subcellular compartment engineering, the technic of oleosome tar-geting had been developed by a research team of Hawaii University at Manoa (USA) for the purpose of designing tunable functional nanoparticles [87]. Surface display of differ-ent heterologous proteins on *Y. lipolytica* LB was obtained through fusion with a heterol-ogous oleosin from plant and the resulting subcellular structure was fractioned through sonication into nano-oleosomes of tunable size. The obtained armed oleosomes corre-spond to stable nanoparticles (200-300 nm of diameter) used as scaffold for protein dis-play, which can be equipped with various functionalities and used for numerous biomed-ical applications such as biosensors, cell targeting or drug delivery [87]. This work has also brought the first proof-of-concept that co-expressing in *Y. lipolytica* different heterol-ogous proteins fused with either a dockerin or a cohesin domain would lead to their *in vivo* self-assembly that, combined with their surface display on oleosomes, can generate functional nanofactories applicable to various biotechnological purposes [87].

Surface display of silicatein on the surface of GM *Y. lipolytica* cells has been used by a research team of Ocean University of China (Qindao) to aggregate armed yeasts into flocs, a sheet-like biosilica-yeast hybrid material, for bioremediation applications [88]. At last, an ambitious and complex engineering work, directed by Huazhong University of Science and Technology (Wuhan, China), led to the design of Euk.cement, an autocementation kit project based on live GM *Y. lipolytica* cells. Euk.cement is expected to stabilize underwater sands by biologically induced autocementation, for civil engineering or environmental restoration purposes. Surface-display and secretion of recombinant peptides and proteins by GM *Y. lipolytica* cells lead firstly to their immobilization onto silica sand particles and secondly to carbonate sedimentation, namely autocementation [89]. The Euk.cement kit project has received a gold medal at iGEM (International Genetically Engineered Machine) Competition in 2015 (MIT, USA - http://2015.igem.org/Team:HUST-China) but is not, to the best of our knowledge, commercialized yet.

2.4. Long-lost relatives: other yeasts of the Yarrowia clade

2.4.1. Brief outline of phylogeny, habitat and characteristics

Y. lipolytica has a pretty chaotic phylogenetic history: in its early days as a yeast species, only the anamorph (i.e. asexual state) was known and classified as *Candida lipolytica*. The teleomorph (i.e. sexual state) was discovered only in 1970 and classified as *Endomycopsis lipolytica*, before being renamed *Saccharomycopsis lipolytica* and, finally, *Yarrowia lipolytica*, as summarized previously [1]. For more than fifty years, *Y. lipolytica* has been considered as the only species of the *Yarrowia* genus and this is only since two decades, when sequencing has become a routine technic for taxonomic studies, that other species have been identified [90-96]. The comparison of sequences from D1/D2 domains of the large subunit (LSU) rRNA gene and from internal transcribed spacer regions (ITS1 and ITS2) has highlighted some heterogeneity among former *Yarrowia* isolates and has also shown that some yeasts previously belonging to the catch-all asexual *Candida* genus needed to be reattributed to the *Yarrowia* clade. Besides *Y. lipolytica*, twelve other *Yarrowia* species are now listed on the NCBI Taxonomy Browser webpage (https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi).

Most of these new species have been isolated in Europe, USA or Asia, from a large range of natural habitats, rather similar to those of *Y. lipolytica*: dairy and meat food products for *Y. alimentaria*, *Y. oslonensis*, *Y. divulgata*, *Y. galli*, *Y. bubula* and *Y. porcina* [1, 90, 91, 95, 96]; mangrove or marine waters for *Y. keelungensis*, *Y. phangngensis* [1, 92, 93] and also for other strains of *Y. divulgata* [95]; guts of insects or vertebrates for *Y. parophoni*, *Y. yakushimensis* and *Y. hollandica* [1, 91, 94]. More unexpectedly, *Y. deformans* was isolated from humans, notably from a fingernail [94], which could raise concerns about its innocuousness (as is also the case for some *Y. galli* isolates [97]. At last, a new species has been proposed more recently for addition to the *Yarrowia* clade, *Y. brassicae*, isolated from a fermented vegetal food, Chinese sauerkraut [98]. These different species of the *Yarrowia* clade have in common to be Crabtree-negative oleaginous yeasts, to grow through multilateral budding, to be able to form both pseudohyphae and true hyphae, to possess an expanded lipase family (7 to 16 genes) and to be physiologically rather similar (e.g. use of a limited number of sugars and polyols as carbon sources) [99]. A few differences were however observed in substrate use: for example, *Y. brassicae* present the unique characteristic of growing on inulin, an interesting asset for potential applications [98]. The maximum growth temperature of these yeasts is rather variable, from 27 °C for *Y. alimentaria* to 37 °C for *Y. phangngensis* [99]. Another oleaginous yeast, *Candida hispaniensis*, has been often associated with an extended version of the *Yarrowia* clade, although this affiliation remained dubious [99]. Published phylogenic trees of yeasts from the *Yarrowia* clade, established from D1/D2 LSU rRNA gene sequences, identify *Y. yakushimensis* as the species closest to *Y. lipolytica*, and *Y. alimentaria* as the furthest one in the *Yarrowia* genus, with *C. hispaniensis* being still further, at a basal position [91, 98, 100]. More recent data, obtained from further study and full genomic sequencing of *C. hispaniensis*, clarified the phylogenic

position of this oleaginous yeast outside from the *Yarrowia* clade, as a close parent of the genus [101]. The term "extended *Yarrowia* clade" will however continue to be used in this review for earlier studies still including *C. hispaniensis*. This phylogenic revision is consistent with the fact that many characteristics of *C. hispaniensis* differentiate this yeast from all the members of the *Yarrowia* clade (no (pseudo)hyphae, compact genome of 10.6 Mb, G/C content of 41%, lipase family reduced to 3 genes) [101]. Nevertheless, *C. hispaniensis* presents some particularly attractive characteristics for industrial biolipid production (30% more oleaginous than *Y. lipolytica* W29 strain, faster replication, no filamentation, no competing citric acid production) and the proof of concept of its amenability to genetic engineering has been very recently demonstrated [101]. A more recent study, focused on variations in the telomeric repeats (encoded in telomerase RNAs - TERs) and in the repeat-binding proteins from all species (except *Y. brassicae*) from the extended *Yarrowia* clade confirmed the homogeneity of the *Yarrowia* genus and the external position of *C. hispaniensis* [102]. This work determined a 10 bp consensus sequence for the TER-derived telomeric repeat unit of all *Yarrowia* species, with minor variations between species, and identified in all of them a functional homologue of *Y. lipolytica* Tay1 protein able to bind all repeat variants. This study provided significant insights into the co-evolution of TERs, telomeric repeats and telomere-binding proteins in the *Yarrowia* clade and, more generally, in yeasts [102].

2.4.2. Potential applications of other yeasts of the *Yarrowia* clade

There has only been a few reports of exploring the biodiversity of the different yeasts from the *Yarrowia* clade for biotechnological applications [99, 100, 103], and a single attempt of engineering one of them with the tools developed for *Y. lipolytica* has been reported in the scientific literature [104]. An INRA study compared growth, lipid synthesis and storage of *Y. lipolytica* and 8 other yeasts from the extended *Yarrowia* clade, during growth on 31 non hydrophobic (sugars, carbohydrates) and 13 hydrophobic (triglycerides, FFA, alkanes) carbon sources. Despite a common oleaginicity, the specific patterns of substrate use and lipid storage varied: for example, the lipid content of cells grown on oleic acid ranged from 30% of CDW for *Y. oslonensis* to 67% for *C. hispaniensis* [103]. In collaboration with the UPWr (Wroclaw, Poland), *Y. lipolytica* and 11 other yeasts from the extended *Yarrowia* clade were tested for erythritol, mannitol or citric acid production from pure glycerol, glucose or fructose. If only *Y. lipolytica* was able to secrete citric acid, several alternative *Yarrowia* strains were able to produce efficiently polyols, including erythritol and mannitol. The best sweetener producer was *Y. oslonensis*, while *Y. hollandica* and *Y. divulgata* could also be promising species, provided optimizing media composition and cultivation parameters [99]. Research teams from the National Center for Agricultural Utilization Research (USDA-ARS, USA) investigated biomass and lipid production capacity of 57 strains of the extended *Yarrowia* clade (45 *Y. lipolytica* isolates and 1-2 isolates of 12 other species) grown on a low-cost renewable feedstock (non-detoxified dilute acid-pretreated switchgrass hydrolysate), in order to expand the diversity of oleaginous yeasts amenable for bioeconomy [100]. With the exception of three of them that were unable to grow in these conditions (*Y. alimentaria*, *Y. yakushimensis* and *C. hispaniensis*), most alternative *Yarrowia* strains were able to accumulate lipids as well or better than the reference W29 *Y. lipolytica* strain, at least for the longer growth times. Two of them, the *Y. phangngensis* and *Y. hollandica* type strains, were top lipid producers with maximum lipid titers respectively fourfold and two fold higher than W29 [100]. A subset of interesting strains has been further characterized for inhibitor tolerance, production kinetics and fatty acid composition, for intended future biotechnological applications.

Interestingly, the same USDA-ARS teams have recently demonstrated that several genetic tools designed for use in *Y. lipolytica* could also be applied to *Y. phangngensis* engineering, with little or no need for modifications [104]. These authors used some promoters, expression vectors and antibiotic resistance genes from *Y. lipolytica*, applied a similar transformation protocol, and finally employed the Cre-lox system for marker recycling designed for this yeast (cf. Section 4.2.5), in order to engineer the PT1–17 type strain of *Y.*

phangngensis for improved lipid production from cellulosic feedstock (acid-pretreated switchgrass hydrolysate). The GM *Y. phangngensis* strain exhibited a 58% decrease in lag time and a 32% increase in lipid titer compared to the parent wild-type strain, due to enhanced detoxification of inhibitor compounds and to added carbon flux into the triacylglycerol synthesis pathway [104]. This demonstration of the easy amenability of an alternative *Yarrowia* strain to genetic engineering is expected to generate a new interest in the highly oleaginous *Y. phangngensis* species and can be hoped to be a good omen for the other yeasts of the *Yarrowia* clade.

3. Fantastic yeasts and where to find them: *Yarrowia* strains and yeast collections

The main purpose of this Section is to provide a useful tool for both advanced users and beginners in *Yarrowia* biotechnology, by reviewing the main *Y. lipolytica* strains of interest with their characteristics and usual domains of application, and by identifying the main yeast collections that makes them (and other species of the *Yarrowia* clade) publicly available worldwide. The Table 1 presents a selection of *Y. lipolytica* wild type or GM strains of industrial/biotechnological interest, a few of which will be described in more detail below. Only the strains that have been deposited in at least one yeast collection have been listed in this Table, which makes them publicly available for research purposes (at the exception of a few of them, as indicated).

Table 1. Selection of remarkable *Yarrowia lipolytica* wild type or engineered (GM) strains of biotechnological interest (haploid, unless specified)

Strain usual name / Origin (Reference number in yeast collections - cf. Table 2)	Genotype Phenotype	Remarkable characteristics	Usual applications
A-101 / carwash effluents, Poland [105] not publicly available (a)	ND wild type prototroph	robust growth on oil [105], high citric acid production [106], sequenced strain [107]	*in situ* soil bioremediation [39, 40], citric acid production [106], UV-mutagenesis and metabolic engineering host for design of improved strains producing citrate or erythritol [54, 108, 109]
ACA-DC 50109 [aka LGAM S(7)1] / Greece [110] not publicly available (b)	ND wild type prototroph	high lipid content and productivity [110]	citric acid and SCO production [110-113], metabolic engineering host for design of improved GM strains [112, 113]
ATCC 18942 [aka YB-423]/ corn-processing plant, USA [114] (CBS 6124, CLIB 183, JCM 2320 & 8060, MUCL 29853, NBRC 1548, NRRL YB-423)	MatA MatB wild type prototroph	diploid type strain [114], robust growth [115]	yeast biomass production [115]

ATCC 20362 [aka 2002]/ USA	ND wild type prototroph	robust growth, high lipid content and productivity [45]	degradation of petroleum crude oil (US patent 3856667A), metabolic engineering host for design of Dupont GM PUFA-producing platform (cf. Section 2.3.2) [45, 46]
ATCC 48436 / soil, Japan [116] (CBS 6303, CLIB 703, JCM 8054, NBRC 10073)	MatA wild type prototroph	produces lipase activators [116]	lipase production [116], parent strain for Artechno highly lipolytic mutants used for bioremediation (cf. Section 2.3.1)
D 1805 / France [117] (ATCC 20390)	MatA MatB wild type prototroph	non-sporulating diploid, robust growth [117, 118], self-cycling fermentation [119]	organic acid production [119]
H222 / soil, Germany [9, 120] (CLIB 80)	MatA wild type prototroph	better fructose assimilation, high citric acid production [120], sequenced strain [26, 29]	organic acid production, metabolic engineering host for design of improved GM strains [121, 122]
NCIM 3589 / marine waters, India [123]	ND wild type prototroph	biofilm formation [124] emulsifier production [125]	gold nanoparticle production [85]
SWJ-1b / marine fish gut, China [4] (MCCC 2E00068)	ND wild type prototroph	high level of crude protein [4]	citric acid and SCP production [126], metabolic engineering host for design of improved GM strains [127, 128] ARTP-mutagenesis host for design of improved strains producing erythritol [129]
W29 / sewage water*, France [10, 130] (ATCC 20460, CBS 7504, CICC 1778, CLIB 89, NBRC 113670, NRLL Y-3178, VKPM Y-3178)	MatA wild type prototroph	high secretion level of proteins [9, 131], sequenced strain [28, 132]	organic acid production [10], basis for the Po1 series of heterologous protein-producing GM strains and the JMY2566 GM strain for high-throughput applications (cf. Fig 2) basis for GM obese strains (cf. Section 3.1.4)
E129 / GM from a W29 and ATCC 18942 crossing [9, 133] (CLIB 121)	MatA, *lys11-23, leu2-270, ura3-302, xpr2-322*	able to grow on sucrose [9, 14] deleted for alkaline extracellular protease [9]	heterologous protein production [131]

(cf. Fig 2)	Lys$^-$, Leu$^-$, Ura$^-$, Suc$^+$, ΔAEP		
E150 / GM from a W29 and ATCC 18942 crossing [9, 133] (CLIB 122) (cf. Fig 2)	MatB, *his1, leu2-270, ura3-302, xpr2-322* His$^-$, Leu$^-$, Ura$^-$, Suc$^+$, ΔAEP	able to grow on sucrose [9, 14] deleted for alkaline extracellular protease [9], reference sequenced strain [25, 26]	reference for assembling and annotating genomes [25, 26]
Po1d / GM from W29 [134] (CLIB 139) (cf. Fig 2)	MatA, *leu2-270, ura3-302, xpr2-322* Leu$^-$, Ura$^-$, Suc$^+$, ΔAEP	able to grow on sucrose [14] deleted for alkaline extracellular protease [134]	heterologous protein production [18, 19, 134], metabolic engineering host for design of GM strains for multiple applications [18, 19] metabolic engineering host for design of Oxyrane ERT-producing platform (cf. Section 2.3.2) [49]
Po1f / GM from W29 [135] (ATCC MYA-2613, CLIB 724, VKPM Y-3155) (cf. Fig 2)	MatA, *leu2-270, ura3-302, xpr2-322, axp1-2* Leu$^-$, Ura$^-$, Suc$^+$, ΔAEP, ΔAXP	able to grow on sucrose [14] deleted for both extracellular proteases [135], sequenced strain [136, 137]	heterologous protein production [18, 19, 135], metabolic engineering host for design of GM strains for multiple applications [18, 19]
Po1g / GM from W29 [135] (CLIB 725) (cf. Fig 2)	MatA, *leu2-270, ura3-302::URA3, xpr2-322, axp1-2* Leu$^-$, Suc$^+$, ΔAEP, ΔAXP	able to grow on sucrose [14] deleted for both extracellular proteases, carry a pBR322 docking platform [135]	heterologous protein production [18, 19, 135], included in the YLEX kit for expression/secretion of heterologous proteins [18, 19] (cf. Section 2.3.2)
Po1h / GM from W29 [17, 35] (CLIB 882) (cf. Fig 2)	MatA, *ura3-302, xpr2-322, axp1-2* Ura$^-$, Suc$^+$, ΔAEP, ΔAXP	able to grow on sucrose [14] deleted for both extracellular proteases [17, 35]	heterologous protein production [17- 19, 35], metabolic engineering host for design of GM strains for multiple applications [18, 19]
Po1t / GM from W29 [17, 35] (CLIB 883) (cf. Fig 2)	MatA, *leu2-270,* LEU2, *ura3-302::URA3, xpr2-322, axp1-2* Suc$^+$, ΔAEP, ΔAXP	able to grow on sucrose [14] deleted for both extracellular proteases, carry a pBR322 docking platform, GM prototroph [17, 35]	negative control for heterologous protein production by other Po1 strains [17, 35]
JMY2566 / GM from W29 [138] (CLIB 1779) (cf. Fig 2)	MatA, *leu2-270, ura3::pTEF-RedStar2-LEU2-zeta, xpr2-322* Ura$^-$, ΔAEP, RedStar2	deleted for alkaline extracellular protease, fluorescent (red) strain,	high-throughput mutant library screening [138]

		carry a zeta LTR sequences docking platform [138]	
WSH-Z06 / oil-polluted soil (refinery), China [139] not publicly available (c)	ND wild type prototroph	thiamine-auxotrophic natural overproducer of KGA [139, 140], sequenced strain [141]	KGA and keto acids production [140], basis for traditionally obtained hyper-producer mutants [141], metabolic engineering host for design of improved GM strains [142, 143]
YB-392 / gluten settler, USA YB-419 / maize fiber tailings, USA (NRRL YB-392 & NRRL YB-419) YB-420, YB-566 & YB-567 (not publicly available, do not appear on online catalog)	ND wild type prototrophs	biomass hydrolysate consumption, inhibitor tolerance, high lipid/fatty acid or sugar alcohol production (cf. Section 3.2.1) [101], sequenced strains [144]	five strains selected as promising candidates for industrial biocatalysis [101, 144]

Strains are listed per alphabetic order of wild-type strain usual name, with their corresponding GM derivatives afterward (by chronological order of construction). (a) deposited at the collection of Wroclaw University of Environmental and Life Sciences. (b) deposited at the Agricultural College of Athens-Dairy Collection. (c) deposited at the China Center for Type Culture Collection (CCTCC M207143). * Erroneously indicated as isolated from soil in other yeast collections. Abbreviations used, per order of occurrence in the table: GM, genetically modified; ND, non-determined or not disclosed; UV, ultraviolet light; SCO, single cell oil; PUFA, polyunsaturated fatty acids; SCP, single cell protein; ARTP, atmospheric and room temperature plasma; ERT, enzyme replacement therapy; KGA, α-ketoglutaric acid.

The different culture collections worldwide that preserve and provide these strains are listed in Table 2, with their website address and their standard of qualification (ISO standard). The total number of strains from *Y. lipolytica* species present in each collection is also indicated, together with that of less identified *Yarrowia* sp. and of other species of the *Yarrowia* clade. This Table is intended as a useful tool to ease the identification, for a given research team, of the closer or more appropriate source for *Yarrowia* strains in general and for precise remarkable strains in particular. For the sake of readability and considering the low number of existing samples of each of the alternative *Yarrowia* species, which makes their search easier, their reference numbers in each collection have not been indicated. The type strain of *Y. phangngensis* that has been the basis for the first engineering assays of alternative *Yarrowia* species described in Section 2.4.2 [104], PT1–17, was deposited at the USA ARS (NRRL Y-63743) and at the Dutch CBS-KNAW (CBS 10407) yeast collections. These two institutions are the more complete sources of alternative species of the extended *Yarrowia* clade in the world: besides *Y. lipolytica*, all 14 other species can be found at CBS-KNAW and 12 of them at ARS (at the exception of *Y. parophoni* and of the more recently described *Y. brassicae*).

Table 2. Main yeast culture collections proposing an important catalogue of publicly available strains of the *Yarrowia* genus or clade

Country (per alphabetic order)	Acronym of the collection (WDCM number)	Full name of the culture collection	Website ISO Standard	Number of strains of the *Yarrowia* genus or clade (*C.* for *Candida*)	Remarkable *Yarrowia lipolytica* strains (haploid, unless specified)
Belgium	BCCM/MUCL (WDCM 308)	Belgian Coordinated Collections of Microorganisms / MUCL Agro-food & Environmental Fungal Collection	http://bccm.belspo.be/about-us/bccm-mucl ISO 9001:2015	29 *Y. lipolytica* + 3 *C. (Y.) alimentaria* 5 *Y. deformans* 1 *C. (Y.) galli* 1 *C. hispaniensis* 2 *C. (Y.) hollandica* 1 *C. (Y.) osloensis** 4 *Y. yakushimensis* (including type strain for all) 2 *Yarrowia* sp.	MUCL 29853: diploid type strain (ATCC 18942)
China (PR)	CICC (WDCM 582)	China Center of Industrial Culture Collection	http://www.china-cicc.org http://www.english.china-cicc.org ISO 9001:2008 ISO 17025:2016; ISO 17034:2018	34 *Y. lipolytica* (with intended biotechnological applications indicated) 1 *Y. brassicae* (type strain)	W29 (CICC 1778) CICC 33063 for erythritol production CICC 31268 & 32291 edible and feed yeasts
	CGMCC (WDCM 550)	China General Microbiological Culture Collection Center	http://www.cgmcc.net http://www.cgmcc.net/english/ ISO 9001:2010; ISO 14001:2010	113 *Y. lipolytica*	
	MCCC (WDCM 1051)	Marine Culture Collection of China	http://www.mccc.org.cn/ ISO 9001:2011	135 *Y. lipolytica* + 1 *Yarrowia* sp.	SWJ-1b (MCCC 2E00068): marine strain, for citric acid and SCP production
France	CIRM-Levures (WDCM 788)	Centre International de Ressources Microbiennes - Levures	https://www6.inrae.fr/cirm/Levures https://www6.inrae.fr/cirm_eng/Yeasts ISO 9001:2015	123 *Y. lipolytica* (including numerous GM laboratory strains) + 4 *Y. deformans* 1 *C. hispaniensis* (including type strain for both)	CLIB 183: diploid type strain (ATCC 18942) CLIB 703: type strain of *Candida paralipolytica*, for lipase production (ATCC 48436)

Country	Collection	Institution	Website / ISO	Species	Notable strains
				2 *Yarrowia* sp.	E122 (CLIB 120): GM E129 (CLIB 121): GM E150 (CLIB 122): GM, sequenced (reference) H222 (CLIB 80): sequenced W29 (CLIB 89): sequenced W29 ura302 (CLIB 141): GM Po1a (CLIB 140): GM Po1d (CLIB 139): GM Po1e (CLIB 723): GM Po1f (CLIB 724): GM, sequenced Po1g (CLIB 725): GM Po1h (CLIB 882): GM Po1t (CLIB 883): GM
Germany	DSMZ (WDCM 274)	Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH	http://www.dsmz.de/ ISO 9001:2008 ISO 17025:2005; ISO Guide 34:2009	6 *Y. lipolytica* + 1 *Y. deformans*	
India	NCIM (WDCM 3)	National Collection of Industrial Microorganisms	https://www.ncl-india.org/files/NCIM/Default.aspx	6 *Y. lipolytica* (erroneously indicated as *Y. lipolitica*)	NCIM 3589: marine strain, for gold nanoparticle production
Japan	NBRC (WDCM 825)	NITE (National Institute of Technology and Evaluation) Biological Resource Center	https://www.nite.go.jp/en/nbrc/index.html https://www.nite.go.jp/nbrc/catalogue/NBRCDispSearchServlet?lang=en ISO 9001:2008	20 *Y. lipolytica* + 1 *C. (Y.) phangngensis* 1 *Y. porcina* (type strain for both)	NBRC 1548: diploid type strain (ATCC 18942) NBRC 10073: type strain of *Candida paralipolytica*, for lipase production (ATCC 48436) W29 (NBRC 113670)
	JCM (WDCM 567)	Japan Collection of Microorganisms	https://jcm.brc.riken.jp/en/ ISO 9001:2015	22 *Y. lipolytica* + 4 *Y. deformans* 1 *Y. keelungensis* 1 *Y. yakushimensis*	JCM 8057: type strain (ATCC 20177) JCM 2320 & 8060: diploid type strain (ATCC 18942)

				(including type strain for all) 5 *Yarrowia* sp.	JCM 8054: type strain of *Candida paralipolytica*, for lipase production (ATCC 48436)
Netherlands	CBS-KNAW (WDCM 133)	CBS Filamentous fungi and Yeast Collection - Westerdijk Fungal Biodiversity Institute	http://www.westerdijkinstitute.nl/ https://wi.knaw.nl/page/Collection https://theyeasts.org/ ISO 9001:2007	38 *Y. lipolytica* + 3 *C. (Y.) alimentaria* 1 *Y. brassicae* 1 *Y. bubula* 18 *Y. deformans* 1 *Y. divulgata* 2 *Y. galli* 2 *C. (Y.) hollandica* 1 *Y. keelungensis* 4 *Y. osloensis* 1 *Y. parophoni** 1 *C. (Y.) phangngensis* 2 *Y. porcina* 4 *Y. yakushimensis* 2 *C. hispaniensis* (including type strain for all)	CBS 8108: type strain (ATCC 20177) CBS 6124: diploid type strain (ATCC 18942) CBS 6303: type strain of *Candida paralipolytica*, for lipase production (ATCC 48436) W29 (CBS 7504)
Russia (Fed)	VKPM (WDCM 588)	Russian National Collection of Industrial Microorganisms	https://vkpm.genetika.ru/	28 *Y. lipolytica*	W29 (Y-3178) Po1f (Y-3155): GM Po1f Ura$^+$ (Y-3483): GM
USA	ATCC (WDCM 1)	American Type Culture Collection	http://www.atcc.org/ ISO 9001:2015 ISO 13485:2016; ISO 17025:2017; ISO 17034:2016	132 *Y. lipolytica* + 1 *Y. deformans* 1 *C. (Y.) phangngensis** (type strain for both)	ATCC 20177: type strain ATCC 18942: diploid type strain ATCC 48436: type strain of *Candida paralipolytica*, for lipase production D 1805 (ATCC 20390): non-sporulating diploid for organic acid production ATCC 20362: basis for Dupont PUFA-producing platform W29 (ATCC 20460)

					Po1f (ATCC MYA-2613): GM
	NRRL (WDCM 97)	Agricultural Research Service (ARS) Culture Collection	http://nrrl.ncaur.usda.gov/	34 *Y. lipolytica* + 1 *C. (Y.) alimentaria* 1 *Y. bubula* 2 *Y. deformans* 1 *Y. divulgata* 1 *C. (Y.) galli* 1 *C. (Y.) hollandica* 1 *Y. keelungensis* 1 *C. (Y.) osloensis* 2 *C. (Y.) phangngensis** 1 *Y. porcina* 1 *Y. yakushimensis* 2 *C. hispaniensis* (including type strain for all)	NRRL YB-423: diploid type strain (ATCC 18942) YB-392, YB-419, YB-420, YB-566 & YB-567: selected as promising candidate strains for industrial biocatalysis, all sequenced W29 (Y-63746)

Data compiled from WDCM (World Data Centre for Microorganisms - www.wfcc.info/ccinfo/home/) and from the collections' websites. Abbreviations used, per order of occurrence in the table: SCP, single cell protein; GM, genetically modified; PUFA, poly-unsaturated fatty acids.*: indicated as their synonyms *oslonensis* on BCCM website, *parophonii* on CBS, or *phangngaensis* on ATCC and NRRL websites.

3.1. Oldies but goodies: elder Y. lipolytica strains and their derivatives

The different *Y. lipolytica* strains described in this Section that are publicly available can be found, for the most widely used of them, in different yeast collections (cf. Tables 1 and 2), but the CIRM-Levures INRAE yeast collection (France) is the only one that can provide all of them.

3.1.1. From Paris sewer to worldwide renown: the success story of W29

The W29 wild type strain has been isolated in the 1970s from sewage water, in Paris (France) [130], and attracted attention for its robustness of growth and high capacity of secretion of various enzymes (proteases, lipase, RNase) [9, 131]. It became rapidly the favourite choice for developing heterologous protein production tools at INRA and was implied in the French *Y. lipolytica* strain inbreeding program [9, 10]. W29 is the most ubiquitous strain in yeast collections worldwide and has become a sort of reference for comparing the performances of new *Y. lipolytica* isolates or of alternative *Yarrowia* species for growth or production of various compounds [100]. The genetic background of W29 has benefited to its various GM derivatives, which constitute since decades the most frequently used strains in the domain of heterologous protein production [17, 18] (cf. Sections 3.1.2 to 3.1.6). This strain has also been applied to organic acid production [10]. Its genome is the third to have been fully sequenced and assembled [28, 132], after those from E150 and Po1f (see below), although these strains were its progeny/derivatives. The mitochondrial sequence of W29 had however been determined earlier [145]. The sequencing strategy used and the *de novo* assembly of its genome have been the first to enable obtaining a single contig for each of the six chromosomes A to F and for the mitochondrial chromosome M [28]. In contrast to the reference sequenced strain E150, W29 is a strain devoid of Ylt1 retrotransposon as well as of solo Ylt1 LTRs (zeta sequences, used as targeting elements in some vectors - cf. Section 4.1.2). Several other genomic structure variations and differences in the presence of retro-elements were observed in the parent W29, compared to the progeny E150, notably the inversion of a 71 kb fragment from the chromosome B

and the presence of a novel Ty3/Gypsy retrotransposon with multiple associated LTR-like sequences [28]. The sequencing and assembly of the wild-type W29 strain also provided data on *Y. lipolytica* rDNA sequences (which were excluded in previous genomic analyses) together with complete sequences for *URA3*, *LEU2*, *HIS1* and *XPR2* genes, which carried deletions in previously sequenced auxotrophic strains [28, 132]. The main GM strains derived from W29 will be described below, with their genealogy represented in Figure 2.
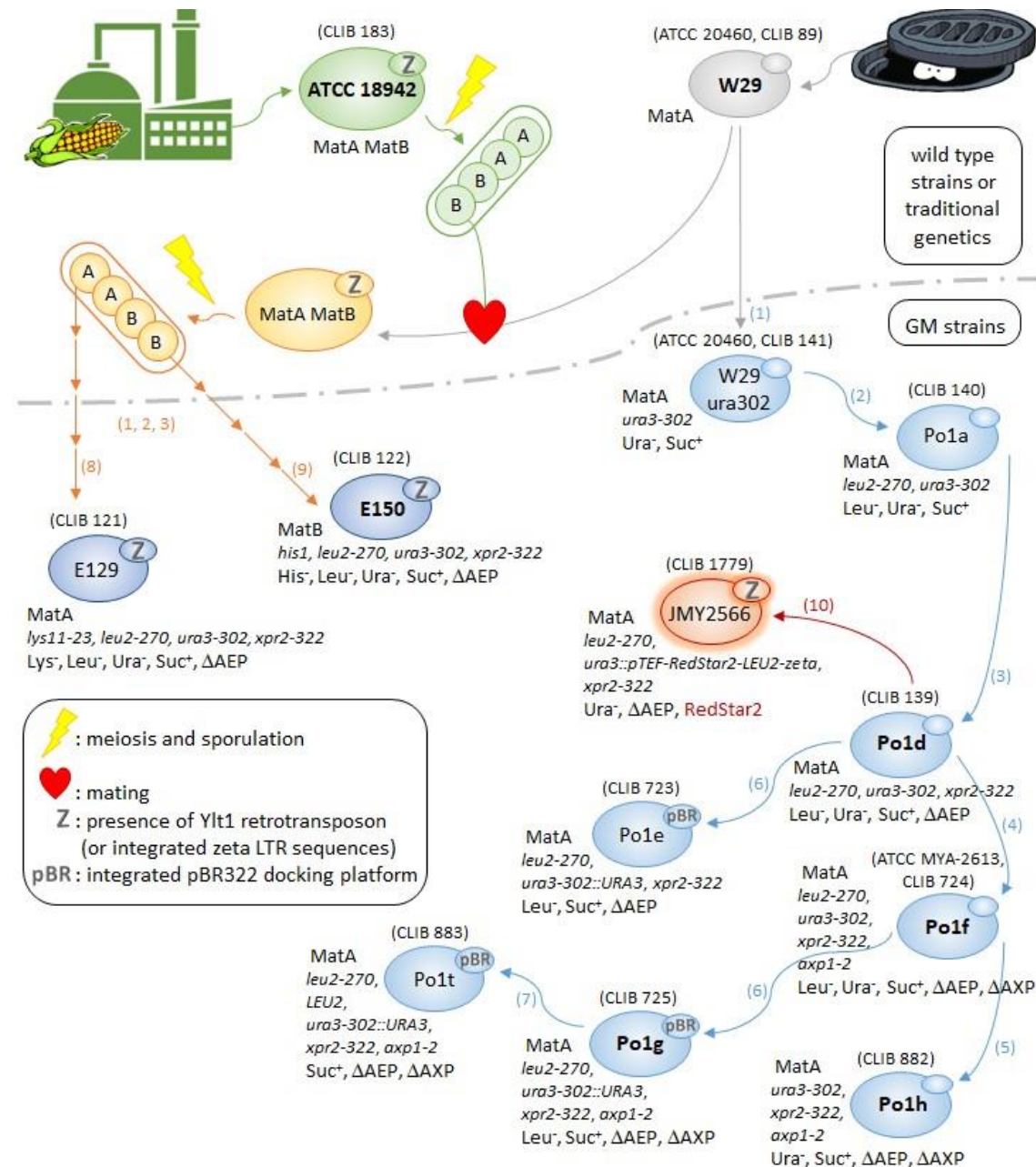

Figure 2. Genealogy of strains of biotechnological interest derived from ATCC 18942 and/or W29
Schematic representation of the origins of ATCC 18942 and W29 wild-type strains and their publicly available derivatives: genotypes and phenotypes are indicated near each strain; yeast collection reference numbers are shown only for ATCC and CLIB-Levures (others can be found in Table 1). Genetic events are represented by the lightning and heart symbols; the presence of sequences that can be used for targeting integrations (zeta sequences, plasmidic backbone) is indicated in the nascent yeast bud. The genetic engineering steps are indicated by bracketed numbers, as follows: (1) genic conversion of *URA3* into *ura3-302* (corresponding to *ura3::pXPR2:SUC2*, namely to a 728 bp *Xho*I-*Eco*RV deletion in *URA3* coding region with disruption by the *SUC2* gene from *S. cerevisiae* under the control of the *XPR2* promoter, conferring the ability to grow on sucrose or molasses) [14];

(2) genic conversion of *LEU2* into *leu2-270* (681 bp *Stu*I deletion in *LEU2* coding region); (3) genic conversion of *XPR2* into *xpr2-322* (149 bp *Apa*I deletion in *XPR2* coding region, inactivating alkaline extracellular protease); (4) genic conversion of *AXP1* into *axp1-2* (655 bp *Nco*I deletion in *AXP1* coding region, inactivating acid extracellular protease); (5) genic conversion of *leu2-270* into *LEU2* (restoration of wild-type allele); (6) integration of *Nco*I-linearized pINA300′ plasmid (*URA3* gene in pBR322 [146]); (7) integration of *Not*I-linearized *LEU2*-carrying pINA1269 plasmid [135] at pBR docking platform; (8) genic conversion of *LYS11* into *lys11-23*; (9) genic conversion of *HIS1* into *his1*; (10) genic conversion of *ura3-302* into *ura3::pTEF-RedStar2-LEU2-zeta* (replacement of the *SUC2* expression cassette disrupting *URA3* with two cassettes, for expressing the RedStar2 fluorescent protein from *TEF* promoter and for *LEU2*, and a zeta LTR sequences docking platform) [138]. The names of the more prominent strains are in bold type. Information gathered from [9, 35] or ATCC and CLIB-Levures websites.

3.1.2. W29 and ATCC 18942 progeny: E129 and E150 strains

The early French/USA *Y. lipolytica* strain inbreeding program aimed at establishing genetic maps and developing new strains through mating of isolates of industrial interest, notably W29 and ATCC 18942, this latter being issued from a corn-processing plant in the USA. The ATCC 18942 strain, sometimes designated as YB-423 from its NRRL reference number, is the *Y. lipolytica* diploid type strain [114], exhibits as such a particularly robust growth and has been notably applied to yeast biomass production [115]. As already said in Section 2.2.2, the mating frequency of two natural Mat-compatible strains is very low and more recent sequencing data, highlighting the surprisingly large difference in the presence of retrotransposons and other retro-elements between strains (see above), could constitute an explanation to this long-dating observation.

Nevertheless, as schematized in Figure 2, the sporulation of the diploid ATCC 18942 strain was performed and a Mat-compatible ascospore was selected and used for mating with W29. The resulting diploid was then subjected to sporulation in order to generate ascospores of both mating types, from which multiple steps of genetic engineering allowed to design the "sister" GM strains E129 and E150, of compatible mating types [9]. The parent ATCC 18942 strain bearing naturally Ylt1 retrotransposon, the derived E129 and E150 strains also contain numerous copies of this retrotransposon, as well as solo zeta sequences, that are available to be used as multiple dispersed targeting elements for further engineering (cf. Section 4). E129 and E150 are both tri-auxotrophic strains that carry the *ura3-302* allele, corresponding to a disruption of *URA3* gene by a heterologous cassette expressing *ScSUC2* that confers them the ability to metabolize sucrose [14]. They are also deleted for the major secreted protease, AEP (alkaline extracellular protease, expressed at neutral/alkaline pHs), encoded by the *XPR2* gene, a characteristic useful for applications in heterologous protein production [9]. The E129 strain has been one of the early strains to be applied to heterologous protein production, but it has been since abandoned in favour of direct derivatives of W29, which secretion system appeared to be more efficient in coping with high increases of secretory pathway cargo load in multicopy-expressing strains [131].

The E150 strain was selected in the French Génolevures II Project [26] to become the first *Y. lipolytica* genome to be fully sequenced, essentially on the basis that, being issued from a crossing between French and USA isolates, it would carry all the different retrotransposons that had already been identified in these *Y. lipolytica* strains at this time. Therefore, E150 now represent the reference strain for the assembly and annotation of *Y. lipolytica* genomes. However, when performing sequence homology searches, it is to be remembered that E150 is a GM strain, carrying deletions in several genes (*URA3*, *LEU2*, *HIS1* and *XPR2*) and containing an added heterologous gene (*SUC2* from *S. cerevisiae*).

Being issued from a meiosis event, the mating type-compatible GM strains E129 and E150 have related, if not identical, genetic backgrounds. Their triple auxotrophy could allow to engineer them using *URA3* and *LEU2* as selection markers and to cross them while selecting for diploid formation on minimal medium, thanks to the remaining auxotrophies for lysine or for histidine, respectively in E129 and in E150. Mating between

these "sister" strains occurs at a satisfactory, if not optimal, frequency (C. Madzak, un-published results) and the resulting diploid could cumulate the genetic modifications and present an improved robustness (cf. Section 4.4.2 for implementation of similar mating strategies).

3.1.3. W29 derivatives: the Po1 series of strains

The favourable genetic background of W29 for protein secretion has prompted the design at INRA of a series of GM derivatives for applications in the domain of heterologous protein production. As represented in Figure 2, W29 has been at first equipped with an *ura3-302* allele (*URA3* disrupted by *ScSUC2* cassette) that provided both an auxotrophy and the ability to use sucrose as sole carbon source [14]. Then, a series of "Po1" GM strains where derived, by addition of leucine auxotrophy, by deletion of the major (AEP) or of both extracellular proteases (AEP and AXP) and finally by complementation of one or of both auxotrophies [17, 18]. The Po1d strain, bearing two auxotrophies due to non-reverting mutations and able to metabolize sucrose [134], remained for more than a decade the most frequently used recipient strain worldwide for heterologous protein production. Po1d is notably the host strain used as chassis in the design of the ERT-producing platform from Oxyrane (Ghent, Belgium) (cf. Section 2.3.2) [49].

The privilege of being the most frequently used host has now been transmitted to the Po1f strain [135], which is additionally deleted for the second *Y. lipolytica* extracellular protease (AXP, expressed at acidic pHs, in contrast to AEP), thus eliminating all secreted proteasic activity that could be a threat for heterologous proteins. Po1f is the second *Y. lipolytica* strain to have had its genome fully sequenced and assembled [136, 137], which reflects the recognition of its predominance as a metabolic engineering host [17-19]. In contrast to the other Po1 strains, which can be obtained only from CIRM-Levures yeast collection, Po1f can also be found at ATCC and VKPM (cf. Tables 1 and 2).

In contrast to Po1d, f and h strains, which can retain the GRAS status of *Y. lipolytica* since their engineering brought only yeast-derived sequences, the strain Po1g has been equipped with an integrated bacterial-derived sequence, namely a pBR322 docking platform, in order to facilitate the further integration of pBR322-based expression vectors. This easy-to-use integration system takes benefit of the large region of homology between these vectors and the docking platform to obtain very high transformation efficiencies (in the range of 10^4 to 10^5 transformants per μg of DNA) and a high percentage of targeted integration (in the range of 80-90%) despite the high level of non-homologous end joining (NHEJ) in *Y. lipolytica* cells [35, 135]. The Po1g strain has been included in the Yeastern YLEX commercial kit (cf. Section 2.3.2), together with two pBR322-based vectors, one for intracellular expression and one for secretion of heterologous proteins. This strain retains only a leucine auxotrophy, allowing its transformants to be prototrophs. The YLEX kit was aimed at easy and rapid testing of heterologous production of a given protein in *Y. lipolytica* and has also been demonstrated to be particularly adapted to enzyme engineering, notably through directed mutagenesis [147, 148]. Namely, targeting to the pBR322 docking platform allows to obtain the integration of a unique copy of the heterologous cassette at a precisely known neutral locus in the genome, so that the effect of the different mutations tested can be directly compared by measuring the recombinant enzymatic activity of the various transformant strains. Interestingly, the Po1g strain has been chosen to constitute the basis of Cell Atlas, a series of seven isogenic strains in which different organelles were rendered fluorescent by tagging with GFP (green fluorescent protein from Aequorea victoria), designed by a consortium of Richland (USA) research teams for the purpose of cell biology applications [149].

Other strains of the Po1 series include Po1h, which retains only an uracile auxotrophy, and the prototrophic Po1t strain, which was designed to serve as a negative control when testing heterologous production of a protein in one of the other Po1 strain. At last, a tri-auxotrophic derivative of Po1f strain has also been recently designed at UT Austin, by addition of a tryptophan auxotrophy [150]. This new strain, named PO1j (PO1f *trp1::loxP*) has not, to the best of our knowledge, been deposited in any public yeast

collection. Similarly, a tri-auxotrophic derivative of Po1d strain, Po1dh, deficient for 5-aminolevulinate synthase, has been constructed to allow the use of an artificial chromosome strategy (cf. Section 4.1.4).

3.1.4. Other derivatives of W29: obese strains

Lipid accumulation in *Y. lipolytica* cells result from a complex pattern of interactions between different metabolic pathways, in different cell compartments: lipid synthesis involves mainly the cytosol, lipid storage the ER and the LB, and lipid mobilization the LB and the peroxisomes. The different strategies used for increasing lipid accumulation in several *Y. lipolytica* strains using metabolic engineering have been extensively reviewed very recently [151] and only a few will be evoked below. The compilation of all previous works indicated not only that W29 and its derivatives have been the most used for this purpose, but also that the W29 genetic background allowed to design the best obese GM strains [151]. Several research teams through the world have used different strategies of genetic engineering to push further the oleaginous potential of W29 derivatives in order to obtain GM strains with up to 75% or even 90% of lipids in their CDW [152-154]. Such obese strains constitute precious tools as powerful chassis that could be applied to the production of various valuable lipid-related products and biofuels.

The strategy chosen at INRA consisted in deleting several genes from the glycerol-3-phosphate (G3P) shuttle pathway and the β-oxidation pathway, in order to increase the availability of G3P and of FFA that both constitute limiting factors for TAG synthesis. This first attempt at improving *Y. lipolytica* oleaginicity allowed to engineer Po1d into an obese strain (Po1d Δ*gut2*Δ*pox1-6*) that can accumulate lipids up to 75% of CDW [152]. More recently, as briefly evoked in Section 2.3.3, the UPWr (Poland) collaborated with INRA, at first to develop some obese strains from either a W29 or an A-101 genetic background, and then to engineer them further for use of a wide substrate range [54]. These teams engineered at first Po1f to generate the Y4086 obese strain (Po1f Δ*pox1–6*, Δ*tgl4*, *pTEF-DGA1*, *pTEF-GPD1*, *pTEF-YlHXK1*, *pTEF-ScSUC2*), with the ability to use sucrose (heterologous expression of *ScSUC2*) and an improved use of fructose (overexpression of hexokinase *HXK1*) [155] and, then, the ability to use other alternative substrates was added through overexpression of the endogenous Leloir pathway (*GAL* genes, for galactose use) [156] and through heterologous expression of the *Kluyveromyces marxianus INU1* inulinase gene (for degradation of inulin into fructose). The resulting YLZ150 obese strain was demonstrated to be able to efficiently use glucose, fructose, galactose, sucrose and inulin for lipid biosynthesis, with the highest lipid concentration (24 g/L) obtained from inulin [54]. The W29 genetic background appeared to be more appropriate for efficient lipid biosynthesis in these experiments than the A-101 one (see more details on A-101-derived obese strains with increase substrate range in Section 3.2.2), with a lipid concentration that was 57% higher [54]. The YLZ150 obese strain, able to use a large range of renewable biomass-derived sugars for SCO or biofuel production, paves the way to a sustainable bioeconomy.

Independently, an ambitious strategy of combinatorial strain engineering from UT Austin allowed an extensive rewiring of the lipid metabolic pathways of Po1f that generated an obese strain with the best lipid level ever reported, of almost 90% of CDW, with a lipid titer of 25 g/L [153]. The UT Austin research team approach combined multiplex inactivation of two genes (in β-oxidation and peroxisome biogenesis pathways) and overexpression of six lipid synthesis target genes, from three distinct metabolic pathways, with phenotypic induction to generate a series of obese strains. Their results allowed to advance fundamental understanding of lipogenesis in *Y. lipolytica*, by demonstrating notably that it can be uncoupled from nitrogen starvation and is dependent on leucine-mediated signalling [153]. The resulting obese strain with the highest performance (Po1f *pex10*, *mfe1*, leucine+, uracil+, *DGA1* overexpression) exhibits a *de novo* lipid accumulation more than 60-fold higher than its parent Po1f strain and has a high industrial potential for SCO and biofuel production [153]. This obese strain was further improved through an evolutionary engineering approach consisting in reiterated steps of random mutagenesis followed by

floating-based selection, taking advance of the increased buoyancy brought to the cells by an enhanced lipogenesis. One of the improved obese strain obtained, E26, exhibits a saturating lipid content of 87% of CDW [157] and was used in the UT Austin strain mating project combining xylose assimilation to valuable compounds production (cf. Sections 2.3.3 and 4.4.2) [58]. Full genome sequencing of the evolved strains revealed a link of the obese phenotype with *uga2* (succinate semialdehyde dehydrogenase) mutations, suggesting an unexpected role of γ-aminobutyric acid assimilation in lipogenesis [157].

At last, a MIT research team from the reported the highest carbon to lipid conversion yield (85% of theoretical maximal) and lipid yield titer (55 g/L) ever obtained from *Y. lipolytica* [154]. This MIT obese YL-ad9 strain has been obtained, from Po1g strain, by simultaneous overexpression of three genes (*ACC1*, *DGA1* and *SCD*), including a δ-9 stearoyl-CoA desaturase gene which product was identified as rate-limiting by reverse engineering from a mammalian cell obese phenotype [154]. Unexpectedly, this YL-ad9 obese strain demonstrated a growth rate threefold higher than its *LEU2*-complemented parent Po1g strain, which explains its productivity and reinforces its competitiveness for developing a robust and effective process for producing biodiesel or other lipid-derived compounds.

3.1.5. Other derivatives of W29: high-throughput expression platforms for protein engineering

The design of efficient enzyme-based industrial processes generally require an optimization of the desired enzymatic properties through protein engineering and/or directed evolution. Some adapted host strains for expression, and high-throughput systems for screening, are required to respond to this technological demand. To fulfil such a purpose, the first *Y. lipolytica* high-throughput expression platform has been developed, nearly fifteen years ago, as the result of a collaboration between INRA, INSA Toulouse and Toulouse University (France). The host strain used in this platform, JMY1212, has a W29 genetic background: it was derived from MTLY60 strain, a Po1d derivative deleted for extracellular lipases (Po1d Δ*lip2*, Δ*lip7*, Δ*lip8*) which was equipped with an integrated zeta docking platform [158]. MTLY60 and JMY1212, both intended for lipid metabolism engineering purposes, have not been deposited, to the best of our knowledge, in any yeast collection. Like explained above for Po1g and its pBR322 docking platform, the zeta docking platform of JMY1212 allows the targeted integration of a unique copy of any zeta-based vector, which facilitates the screening of new or improved enzymatic functions [158]. In addition, the fact that zeta-based plasmids are auto-cloning vectors (integration of a cassette devoid of bacterial sequences - cf. Section 4.1.2) can allow to retain the GRAS status of *Y. lipolytica*. The JMY1212 high-throughput expression platform was notably applied to engineering *Candida antarctica* lipase B (CalB), a widely used enzyme in industrial biocatalysis, for which it allowed the construction of large libraries of CalB mutants and the screening of new variants with a higher catalytic efficiency [159].

More recently, another high-throughput expression platform has been designed at INRA, based on the JMY2566 strain (also a Po1d derivative), as described in Figure 2 and in Table 1. JMY2566 retains only an uracile auxotrophy and, like JMY1212, is equipped with an integrated zeta docking platform that eases both its transformation with any zeta-based vector and the ulterior screening of enzymatic properties [138]. In contrast to JMY1212, JMY2566 can secrete lipases but is deleted for the extracellular AEP protease. In addition, JMY2566 also bears a cassette for constitutive expression of a red fluorescent protein, RedStar2, which can be used as reporter for cell growth in non-translucent media, during the screening steps [138]. The JMY2566 recipient strain can notably be used as host for the Gateway overexpression vector JMP1529, using an optimized high-throughput transformation method in 96-well plates, so that the platform can be applied to high-throughput mutant library screening. Alternatively, this platform can also be applied to functional exploration of *Y. lipolytica* genetic regulation, as demonstrated by the recent analysis of 148 putative transcription factors by a systematic overexpression approach [160].

3.1.6. Other derivatives of W29: glyco-engineered strains for producing therapeutic proteins

The differences existing between the N-glycosylation pathways in yeasts and in mammalian cells can pose problem for the production of therapeutic proteins. Namely, glycoproteins produced in yeasts display high mannose-type N-glycans that could reduce the *in vivo* half-life of proteins and render them immunogenic in humans and other mammals. To palliate this problem, numerous research teams throughout the world have developed glyco-engineered (aka humanized) strains in the yeast species currently applied to heterologous production [21]. Concerning *Y. lipolytica*, a consortium of Belgian laboratories from VIB (Vlaams Instituut voor Biotechnologie) and Ghent University, together with Oxyrane (Ghent, Belgium), have constructed several glyco-engineered strains able to produce more human-compatible glycoproteins, from W29 and one of its derivatives constructed at INRA, MTLY60 (cf. Section 3.1.5). Notably, glyco-engineered strains able to produce glycoproteins homogeneously carrying Man5GlcNAc2 residues were obtained by the deletion of a two yeast-specific mannosyltransferases and heterologous expression of a fungal mannosidase [20]. Similarly, the mannosyltransferase deleted strain was further engineered by overexpression of a glucosyltransferase and heterologous overexpression of a fungal mannosidase and a fungal glucosidase, which generated a strain able to produce glycoproteins homogeneously carrying Man3GlcNAc2 residues, a core structure common to the different mammalian N-glycans that can additionally be modified in vitro to generate any kind of complex-type N-glycan [161]. Besides the Oxyrane engineering platform able to add M6P residues to glycoproteins, for the development of ERTs for lysosomal storage diseases [49] described in Section 2.3.2, these glyco-engineered strains could contribute to upgrade *Y. lipolytica* into a valuable source of recombinant proteins carrying humanized N-glycans structures compatible with therapeutic applications.

3.1.7. The outsider H222

The H222 wild type strain has been isolated in the 1980s from soil samples, in Leipzig (Germany), and attracted attention for its robust growth on fructose as sole carbon source, markedly better that those of other *Y. lipolytica* strains, and its high level of citric acid production [9, 120]. H222 has been mainly applied to organic acid production and has been genetically engineered at the Technische Universität Dresden (TUD, Germany), in collaboration with INRA, to design GM strains with improved citric acid production and able to metabolize sucrose (with the *ura3-302* allele bringing a *ScSUC2* cassette) [121, 122]. TUD also collaborated with Helmholtz-Zentrum für Umweltforschung (UFZ, Leipzig, Germany) to engineer H222 for improved production of succinic acid or KGA [162, 163]. The robust growth of H222 on fructose was demonstrated to be due to a better fructose assimilation, compared notably to W29 [155] and a study of *Y. lipolytica* sugar transporters explained further the superiority of the H222 strain in this domain [164]. Notably, one of the functional fructose transporter of H222 was found to exist only as a silent pseudogene in W29, and there was some amino acid polymorphism in eight hexose transporters between the two strains, which could account for the better fructose uptake of H222 [51]. Although the subject of previous genomic studies [25, 26], the H222 strain has been fully sequenced and assembled only recently [29]. This *de novo* genomic assembly of H222 revealed important differences with the reference strain E150, notably three major chromosomal rearrangements (two reciprocal translocations and a 300 kb inversion) that explain the observed variations in chromosome sizes [29]. This work illustrates the importance of preferring *de novo* genomic assembly over reference-assisted scaffolding for *Y. lipolytica* and highlights the important chromosomal rearrangements that can be found between strains.

3.2. New kids on the block: Y. lipolytica strains isolated or noticed more recently

3.2.1. Gold diggers: how to find nuggets in old mines

The yeast collections listed in Table 2 constitute a valuable resource of biodiversity in the search for new *Y. lipolytica* genetic backgrounds and some of them have been noticed to be particularly adapted for peculiar applications, as shown in Table 1. For example, as said above in Section 2.3, two wild-type isolates from the US ATCC collection, ATCC 20362 and ATCC 48436, have been selected for commercial applications. The strain ATCC 20362, also referred to as 2002 strain in some publications, is an USA isolate that was cited in the US patent 3856667A from Bioteknika International Inc., in the 1970s, for applications in the degradation of petroleum crude oil. It was selected much more recently by a Dupont research team, for its robustness and oleaginicity, and applied to the design of their GM PUFA-producing platform (cf. Section 2.3.2) [45, 46]. The strain ATCC 48436, an isolate from Japanese soil samples, has been noticed in the 1960s for its high levels of lipase activators and lipase activity [116], which prompted its more recent selection by INRA for the development of highly lipolytic mutants [41], currently applied to bioremediation by Artechno (cf. Section 2.3.1). Similarly, a tropical marine strain from the Indian NCIM collection, NCIM 3589, has been studied for hydrocarbon degradation, biofilm formation and emulsifier production [123-125] before being used for the production of gold nanoparticles [86], which have applications notably in medicine (antibacterial and anticancer properties), in some diagnosis tests and in biomedical imaging.

At last, a consortium of USA laboratories from the National Center for Agricultural Utilization Research (USDA-ARS, USA) performed a wide screening of the biomass and lipid production capacity of 45 *Y. lipolytica* wild-type isolates from the USA NRRL (ARS) collection under industrial type conditions, namely using a non-detoxified dilute acid-pretreated switchgrass hydrolysate as low-cost biomass feedstock [100]. This study identified five promising candidate strains, NRRL YB-392, YB-419, YB-420, YB-566, and YB-567, which genetic backgrounds are well adapted for industrial use, demonstrating better biomass hydrolysate consumption, inhibitor tolerance and lipid production than the W29 reference strain under those conditions. Notably, the three best strains (YB-392, YB-419 and YB-420, from which only the first two appear to be publicly available) produced from 43 to 64% more lipids than the W29 reference strain. These three strains can grow efficiently on 90% undetoxified switchgrass hydrolysate and convert cellulosic sugars into SCO with a fatty acid profile similar to rapeseed (canola), hence suitable for biodiesel production. Two other strains (YB-566 and YB-567, not publicly available) can assimilate xylose more efficiently than usual *Y. lipolytica* strains and convert cellulosic biomass into sugar alcohols like xylitol or arabitol [100]. This work has also contributed to revise some inexact notions about pentose assimilation in *Y. lipolytica*: the fact that this yeast cannot use xylose as sole carbon source does not imply that its pentose assimilation pathway is missing or incomplete. In fact, as reviewed in their publication, xylose (and arabinose) consumption following glucose depletion was reported and bioinformatics tools demonstrated that *Y. lipolytica* has a functional xylose pathway, which enzymes are poorly expressed and for which xylitol dehydrogenase constitutes a limiting step [56]. Among the 28 *Y. lipolytica* strains (on the 45 tested) that were able to exhaust glucose in the hydrolysate, there was a surprisingly large variability (of a tenfold range) in the ensuing xylose consumption [100], which illustrates the interest of exploring strain biodiversity when searching for peculiar capacities. The five strains with high potential for industrial biocatalysis selected in this study have been recently sequenced [144], with the aim to allow further improvement of their robust metabolism by activating and/or rewiring pathways for improved complex sugar assimilation and lipid accumulation.

3.2.2. Finders keepers: for a good isolate, help yourself

Some research teams have preferred to developed applications from wild-type *Y. lipolytica* strains they have themselves isolated, generally from oil-polluted soils or from marine environments. In the 1990s, a research team from the Academy of Agriculture of Wroclaw (Poland) isolated the strain A-101 from oil-polluted soil samples from a car wash site, obtained some UV-generated mutants from it, and compared these strains to a selection of ATCC *Y. lipolytica* strains. A-101 exhibited a robust growth on oil and a high citric

acid production, which was enhanced in some mutant derivatives [105, 106]. As described in Section 2.3.1, A-101 was also applied with success to *in situ* soil bioremediation [39, 40]. This strain was deposited at the collection of the UPWr (Wroclaw, Poland) but is not publicly available. More recently, A-101 has become a metabolic engineering host for the design of improved strains producing citrate or erythritol from various substrates, in the context of collaborations between UPWr and INRA [54, 108, 109]. Notably, some derivatives of this strain were engineered for highly efficient sucrose consumption throught an optimized expression/secretion of the heterologous *ScSUC2* invertase [108]. A draft genome sequence of A-101 has been made available, which interestingly revealed that this strain carries, among multiple retro-elements, a single copy of a solo zeta sequence (LTR of Ylt1) [107]. At last, as briefly evoked in Section 2.3.3, UPWr and INRA developed an obese strain from A-101 and engineer it for using a wide range of substrate [54]. In order to minimize the number of genetic modifications required in this strain, the β-oxidation pathway was blocked by deleting the multifunctional enzyme gene *MFE2*. In order to improve TAG biosynthesis, glycerol-3-phosphate dehydrogenase (*GPD1*) and diacylglycerol acyltransferase (*DGA1*) genes were overexpressed. In order to improve fructose use, a hexokinase (*HXK1*) and an efficient hexose transporter (*YHT3*) were also overexpressed. At last, in order to allow inulin use, a heterologous *K. marxianus* inulinase gene was added [54]. The resulting Y4779 strain was specifically tailored for a better use of fructose and did not compared very favourably with obese strains derived from a W29 background (cf. Section 3.1.4), but it appeared to be particularly adapted to the development of sustainable processes for citric acid or polyhydroxy alcohol production [54].

The strain ACA-DC 50109, also referred to as LGAM S(7)1 strain in some publications, has been isolated at the Agricultural University of Athens (Greece) and applied to citric acid and SCO production, thanks to its high lipid content and productivity [110, 111]. This strain has been deposited at the Agricultural College of Athens-Dairy Collection but is not publicly available. Genetic engineering of ACA-DC 50109 has been initiated at the Ocean University of China (OUC, Qingdao), after an uracil mutant of this strain was isolated using counter-selection on 5'-FOA (5'-fluororotic acid, a toxic uracil precursor), and this strain is used as metabolic engineering host for the design of improved GM strains, notably for SCO production from alternative substrates such as inuline [112, 113]. Interestingly, a global analysis of the acetylproteome (lysine acetylation of proteins, a major post-translational modification) has been performed in this strain, at the OUC, which revealed that lysine acetylation sites were present in more than 22% of *Y. lipolytica* proteins [165]. Notably, lysine acetylation was demonstrated in 65 enzymes from the lipid biosynthesis pathways, which highlights the crucial role that reversible acetylation could play in *Y. lipolytica* and possibly in other oleaginous microorganisms [165].

The Unesco Chinese Center of Marine Biotechnology, at OUC (Qingdao, China), is an institution devoted at exploiting marine microorganism resources for white biotechnology. In the 2000s, their research team isolated 78 wild-type *Y. lipolytica* strains from various marine-related environments (seawater, algae, marine or saltern sediments, guts from marine fishes), which were screened for interesting properties, notably for crude protein content. Seven of these marine strains were shown to contain more than 41% of proteins per CDW, with SWJ-1b having the highest crude protein level [4]. This strain, isolated from the guts of a marine fish, was applied to SCP and citric acid production [126] and became a metabolic engineering host for the design of improved GM strains [127], notably able to metabolize inulin [128]. SWJ-1b Can be obtained from the Marine Culture Collection of China (MCCC 2E00068). More recently, non-GM mutant strains were generated from SWJ-1b by a consortium of laboratories from Huaiyin Normal University and Institute of Technology (Huaian, China) [129]. These research teams used the recent atmospheric and room temperature plasma (ARTP) mutagenesis method, which constitutes a safer and more efficient alternative to traditional technics (e.g. UV, mutagens) allowing to generate mutations in the genome. One of these mutant strain, M53, exhibited the highest erythritol yield (0.65 g/g, from glycerol, in fed-batch fermentor) ever reported for *Y.*

lipolytica or other microorganisms, which demonstrates its potential for erythritol production on a commercial scale [129].

The strain WSH-Z06 has been isolated rather recently, a decade ago, in the screening of 100 oil-polluted soil samples from 20 oil refineries from Wuxi (China) for KGA-producing yeasts, at Jiangnan University (Wuxi, China) [139]. This *Y. lipolytica* strain is auxotrophic for thiamine and a natural overproducer of KGA [139, 140]. It has been deposited at the China Center for Type Culture Collection but is not publicly available. The WSH-Z06 strain has been applied to the production of KGA and of other keto acids [140]. This wild-type isolate has been sequenced and has been used as basis for a complex mutagenesis strategy (separate or combined use of ARTP with UV and mutagens) that generated a series of five hyper-producer mutant strains that were also fully sequenced [141]. This comparative genomics analysis provided insight into the physiology of KGA accumulation in *Y. lipolytica*, notably by highlighting a positive correlation with mitochondrial biogenesis and energy metabolism [141]. Interestingly, a proteomic analysis of the response of this strain to an acidic pH stimuli allowed a better understanding of the metabolic flux shift to KGA observed at lower pH, to be linked to its antioxidant role, and suggested process optimization strategies for short-chain carboxylate production [166]. At last, WSH-Z06 strain is also a metabolic engineering host for the design of GM strains improved for KGA production [142, 143].

4. A brave new world of engineered strains: tools and strategies for building *Y. lipolytica* cell factories

The process of transforming a selected *Y. lipolytica* host strain into a successful cell factory represents a long journey, through multiple technical steps requiring complementary expertises, that is tentatively schematized in Figure 3. As represented in this Figure, remodelling the metabolic pathways of *Y. lipolytica* for the production of a compound of interest can be obtained via deletion/repression/activation/overexpression of endogenous genes combined with (over)expression of a few heterologous genes as well as introduction of complete new metabolic pathways, all steps achieved through classical or more recently developed engineering/editing methods. The data gathered from different omics technologies could allow to develop genome-scale metabolic models that could influence the next engineering steps in a virtuous circle. Adaptative evolution strategies could also be applied for further improvement and, at last, bioprocess engineering will allow to valorise the laboratory achievements into an industrial-scale economically viable bioprocess.
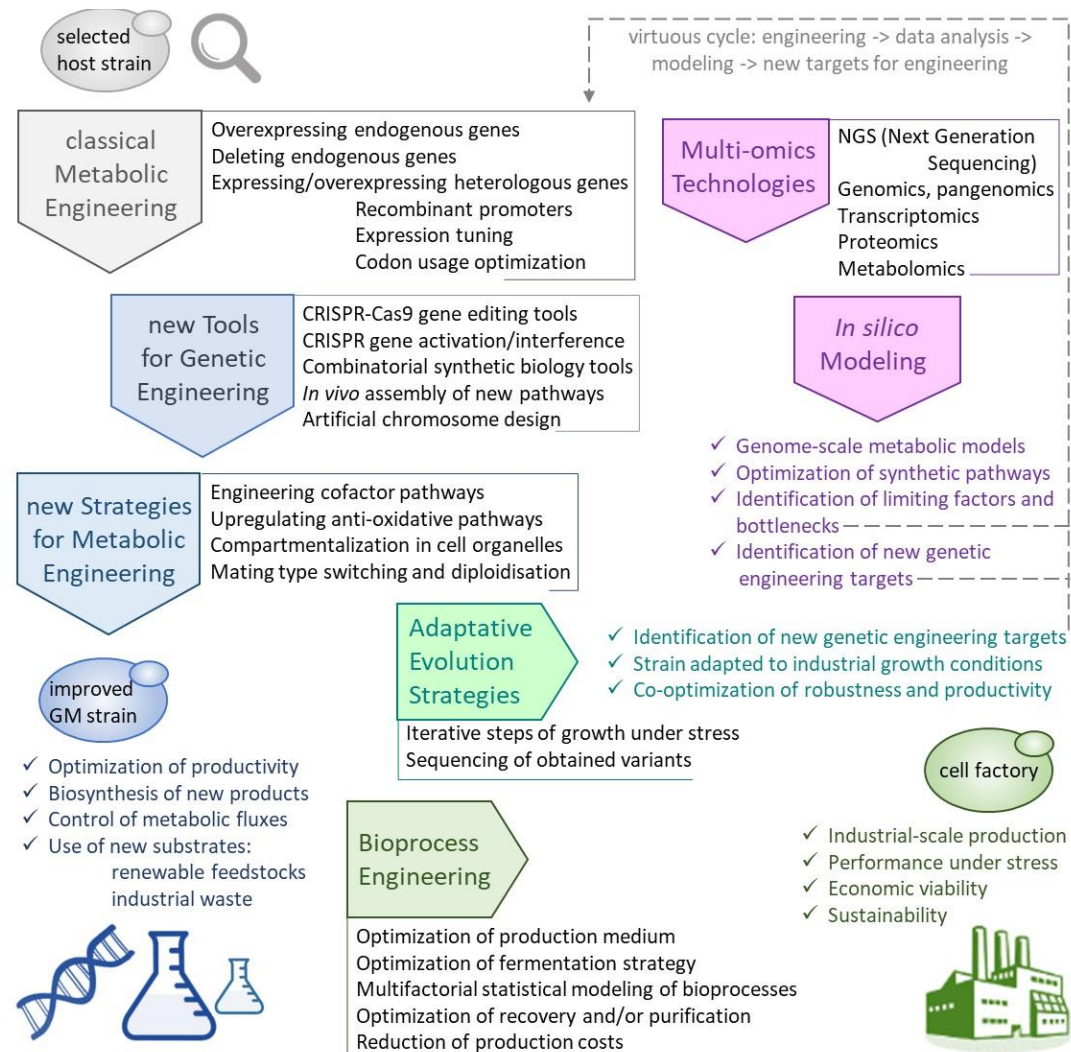
Figure 3. Schematic representation of the possible steps and available tools for engineering a *Y. lipolytica* strain into an effective cell factory

As explained in more detail in the text, both classical metabolic engineering methods and more recently developed tools, such as CRISPR-derived strategies, can be applied to remodelling metabolic pathways for production of a compound of interest. In addition, new strategies for metabolic engineering take also into account the availability of cofactors, the reduction of oxidative compounds and the compartmentalization of the modifications in different cell organelles, in a holistic view of the metabolic fluxes. The obtained GM (genetically modified) strain can also benefit of multi-omics technologies which, by allowing *in silico* modelling of genome-scale metabolic pathways, could contribute to identify limiting factors and bottlenecks, suggesting future genetic engineering targets in a virtuous circle. Adaptive evolution strategies could be employed to better adapt the GM strains to the stress of industrial growth conditions. At last, bioprocess engineering is required to develop viable and sustainable cell factories.

Most of these processing steps have already been extensively reviewed previously [18, 19, 42, 43, 167-172] and this would be out of the scope of this publication to detail all of them, but a brief overview of the available genetic engineering tools will be presented, with a focus of those that could allow to leverage the natural biodiversity of *Y. lipolytica* strains and, possibly, of other yeasts from the *Yarrowia* clade.

4.1. To be or not to be integrated: types of vectors and assembly methods

Metabolic engineering of *Y. lipolytica* requires the assembly of single or multiple expression cassettes (or transcription units, TUs), each composed of a promoter, of an open reading frame (ORF) and of a terminator. In addition to the homologous or heterologous

gene of interest, targeting components can be optionally included into the ORF, in order to direct the resulting recombinant protein to precise intracellular organelles or to the secretion pathway, either for release into the cultivation medium (vesicular secretion) or for display on the cellular surface (surface display). When classical genetic engineering strategies are used, TUs are carried by either integrative or replicative shuttle vectors, built and propagated in *Escherichia coli* strains, that are then introduced into *Y. lipolytica* cells rendered competent using chemical treatments [35, 138, 173] or electroporation [174, 175].

The co-expression of several heterologous genes is generally needed when the introduction of entire metabolic pathways is required to confer new potentialities to *Y. lipolytica* strains. Vectors carrying multiple TUs can be used for that purpose, as reviewed previously [42, 43] and resumed below. If some ambitious genetic engineering works have been performed using the classical (laborious and time-consuming) method of sequential integration [45], complex engineering projects can now benefit of new rapid *in vitro* or *in vivo* DNA assembly methods that were extensively reviewed previously [42] and will only be briefly evoked below in Sections 4.1.3 and 4.1.4.

4.1.1. Episomal vectors

In order to be able to replicate in *Y. lipolytica*, an episomal vector needs to carry an ARS (autonomously replicating sequence)/CEN sequence, bearing co-localized replicative and centromeric functions. Such replicative vectors behave like mini-chromosomes, present in only one or a few copies per cell and easily lost during mitosis [176, 177], which limit their interest for heterologous protein production. In contrast, they constitute the preferred tool for transient expression, as required for marker rescue using the Cre-lox system [178] or for gene editing using CRISPR tools (cf. Sections 4.2.5 and 4.3.1). Replicative vectors have also been chosen for developing some of the recently designed genetic toolkits for *Y. lipolytica* pathway engineering, like the YaliBricks system [179, 180] (cf. Section 4.1.3). Interestingly, some engineering of a *Y. lipolytica* centromeric region by fusion with upstream promoters was performed at UT Austin and allowed to increase both copy numbers and expression levels of the resulting replicative vectors [181], which may possibly give a new impulse to the use of such episomal vectors. Interestingly, the multipurpose pYL15 vector that was used to establish Cell Atlas (cf. Section 3.1.3) by enabling expression of heterologous genes into recombinant fluorescent fusion proteins by addition of a GFP tag, was designed to be used either as a centromeric replicative vector or, following restriction digest, as an integrative one targeting the *leu2* locus of Po1g strain [149].

4.1.2. Integrative vectors and cassettes

Since the beginnings of *Y. lipolytica* genetic engineering, integrative vectors constitute the preferred tool for either heterologous protein production or metabolic pathway engineering, thanks to the very high stability of integrated TUs, comparable to that of native genes [9, 18]. Integration of complete vectors can be targeted to a precise locus in the genome through linearization in a region of homology that, because of the predominance of NHEJ recombination in *Y. lipolytica*, needs to be large enough (0.5–1 kb on each side) [9, 178]. Such regions of homology with the genome that allow the targeting of a vector (or of a TU-bearing DNA fragment generated by restriction digest or PCR) are generally chosen in rDNA coding sequences [134, 182], in the 3' non-coding regions of some genes or in the ORF of selected genes which inactivation could be beneficial for the intended purpose (as proposed notably in the modular Golden Gate toolkit developed at INRA [183]).

A popular alternative is the prior construction of a recipient strain bearing an integrated docking platform, which can be for example a bacterial vector backbone (like the pBR322 docking platform of Po1g strain [135] - cf. Section 3.1.3) or a yeast-derived sequence naturally absent from this strain, like the zeta docking platform of JMY1212 or JMY2566 [138, 158] (cf. Section 3.1.5). As explained in these previous Sections, the docking platforms of these strains allow the easy targeted integration of unique copies of the corresponding vectors at precisely known neutral loci in the genome, facilitating notably enzymatic screening or TU design comparison. The Po1g strain was constructed for easy

integration of the pYLEX1 (aka pINA1269 [135]) expression vector or the pYLSC1 (aka pINA1296 [135]) secretion vector, both included in the YLEX kit (cf. Section 2.3.2). The JMY1212 and JMY2566 strains were designed for efficient integration of any of the numerous available zeta-based vectors [34, 35, 182, 184] or, more precisely, of the TU-bearing transformation cassette that needs to be isolated using restriction digest from these auto-cloning vectors. The zeta-based vectors have been developed at INRA as auto-cloning vectors from which a *URA3*-bearing transformation cassette can be isolated and used to transform any *Y. lipolytica* Ura⁻ strain. These zeta-bordered transformation cassettes can integrate by homologous recombination into the zeta sequences that the strain may bear, for Ylt1-carrying strains (or strains equipped with a zeta platform), or can integrate at random sites in strains devoid of this retrotransposon, thanks to the high level of NHEJ from *Y. lipolytica* [48, 182].

Some zeta-based vectors bear a defective version of the *URA3* marker (*ura3d4* allele with truncated promoter) in order to promote an *in vivo* amplification of the copy number of the integrated transformation cassettes [48, 134, 182]. Despite being a laborious and rather unreliable process, the construction of multicopy strains using this strategy has been fruitfully applied to the production of numerous heterologous proteins, as extensively reviewed previously [18]. These examples contributed to the success of the zeta-based vectors that remained until recently the most widely used tools for engineering *Y. lipolytica*. As a method for amplifying expression, the defective zeta-based vectors now show their limitations: their transformation efficiency is low (dramatically low if random integration is used) while they require testing a high number of transformants to select a good producer. Some promoter engineering strategies, improved during the last decade [185], are now able to drive better expression increases than the use of defective markers, while being more reliable and easier to perform technically (cf. Section 4.2.1).

Using a random integration strategy also present some important drawbacks, notably a lack of control of the construction and a high risk of deleterious effects of the genomic insertions, which are not desirable for industrial production. A precise comparison of random and targeted integration has been performed recently: following the assembly of a β-carotene metabolic pathway in a zeta-based vector, using a Golden Gate Assembly method, the resulting transformation cassette was either integrated at random into the Po1d strain or targeted to the zeta docking platform of the JMY1212 strain. The randomly obtained transformants exemplified highly variable levels of carotenoid production, while the targeted ones showed more reproducibly high yields [186]. Similarly, some instability problems were observed when constructing strains with randomly integrated multiple copies, a factor that impairs the use of such strains for industrial applications, as reviewed recently [43]. For these reasons, the use of targeted integration at selected genomic loci or docking platforms is now considered as a much more reliable strategy.

4.1.3. Multiple transcription unit vectors, *in vitro* DNA assembly methods and *Y. lipolytica* toolboxes

Following the first example of a tandem dual cassette vector, used for engineering Po1g strain for gamma-linolenic acid (GLA) production [187], multiple TUs vectors have been designed in different laboratories. Notably, an integrative vector carrying three TUs was applied to engineering *Y. lipolytica* for improved glycerol metabolism and the resulting strain demonstrated a high stability of the triple integrated expression cassette, despite the presence of tandemly repeated promoter and terminator sequences [188]. The record of TU number and overall vector size belongs to a very large (19 kb) replicative vector carrying 5 TUs that was applied to engineering Po1g strain for violacein biosynthesis [179]. Despite some lack of evidence for estimating the risk of reshuffling problems when using repeated sequences, the research teams who developed toolboxes for *Y. lipolytica* engineering tend to privilege the use of different promoters and terminators as multiple TU elements in order to minimize the possibilities for homologous recombination. Interestingly, a recent process development study using a starch-utilizing GM *Y. lipolytica* for

lipid production suggested that, beside TU design itself, the positional order of multiple TUs was an additional factor to be considered for optimized results [189].

The violacein pathway construction was obtained using the YaliBricks system, an *in vitro* assembly method that allowed the assembly of the 12 kb five-gene violacein biosynthetic pathway in only one week [179]. The YaliBricks system consists in a modular assembly method designed at the University of Maryland (Baltimore County) that complies with the BioBrick standards by using four distinct but ligation-compatible restriction sites for allowing the one-step assembly of several DNA fragments. A library of YaliBricks vectors, carrying various elements for assembly of *Y. lipolytica* TUs (including 12 different promoters) was designed [179], further completed with CRISPR-Cas9 tools [180] (cf. Section 4.3) and all these plasmids are, to the best of our knowledge, intended to be deposited on the Addgene nonprofit repository platform (www.addgene.org).

A different *in vitro* assembly method is proposed in the multipurpose Easy-CloneYALI toolbox developed at the Novo Nordisk Foundation Center for Biosustainability (Technical University of Denmark): their expression toolkit comprises a series of 26 auto-cloning vectors compatible with USER (uracil-specific excision reaction) cloning, a method employing an USER enzyme to create single-stranded extensions in PCR fragments allowing directional seamless assembly of DNA fragments [190]. These Easy-CloneYALI vectors are designed to target the integration of TUs at 11 intergenic sites chosen to allow both an unaffected growth of the recipient cell and a high expression level of the gene of interest; they are available from Addgene [190]. A second toolkit, for CRISPR-Cas9 tools, was also developed (cf. Section 4.3.1).

At last, the Golden Gate Assembly (GGA) method, using Type IIs restriction enzymes (cutting DNA at distance from the recognition site) to generate DNA fragments with variable non-palindromic overhangs that allow their one-step directional assembly, has been chosen at INRA for developing a modular Golden Gate *Y. lipolytica* expression toolkit [183, 191]. The resulting toolbox, available from Addgene, comprises an auto-cloning zeta-based destination vector backbone and a series of 64 plasmids, bearing as many different GG biobricks (including nine promoters) for one-step GGA of auto-cloning expression vectors carrying one, two or three TUs. The resulting integration cassettes can be inserted at random in the genome or targeted to four selected loci, a zeta docking platform and three ORFs from genes which inactivation could be beneficial for industrial applications, namely *LIP2*, *GSY1* and *MFE* [183]. This system was applied notably to optimizing expression from a β-carotene biosynthesis pathways using a promoter shuffling approach. Subsequent integration of this optimized GG-assembled pathway into a Po1d-derived obese *Y. lipolytica* strain, in two copies, generated the best β-carotene yield ever obtained from a microbial producer [192]. This GGA platform was also successfully applied to the assembly of a three-gene metabolic pathway allowing xylose utilization by *Y. lipolytica* [183]. More recently, some CRISPR-Cas9 tools were also added to this GGA toolbox (cf. Section 4.3.1). Very recently, the modular and combinatorial properties of the GGA method has been applied to optimizing violacein production in a collaborative work between the University of Maryland and Jiangnan University. These authors constructed a library of violacein producing GM Po1g strains in which each of the five genes from the violacein pathway could be expressed from any of three promoters of different strengths [66]. The GGA efficiency was maximized by selecting the more effective linker sequences. An optimization of the cultivation conditions of the selected best producer, using the strongest promoter for all genes, lead to the unprecedented yield of 70 mg/L of violacein [66]. This work constitutes a proof of concept of how the construction of GG-assembled pathway libraries with randomized promoter strengths could maximize the output yield from any given heterologous (or possibly native) metabolic pathway.

4.1.4. *In vivo* DNA assembly methods by homologous recombination and artificial chromosomes

Despite the high level of NHEJ in *Y. lipolytica*, a few research groups have attempted rather recently to use homologous recombination (HR) for *in vivo* assembly of metabolic

pathways in this yeast. A consortium of Shanghai laboratories was able to obtain the assembly and integration at the rDNA locus of a *Y. lipolytica* strain of an 11 kb and three-gene β-carotene synthesis pathway in only one week [193]. The whole process included two steps: first, the separate *in vitro* assembly of each of the 3 TUs from the new pathway and of the one for the selection marker, by overlap extension PCR and, then, the co-transformation of yeast cells with these four elements that were directionally assembled and integrated, by *in vivo* HR between their respective overlaps. The overall efficiency of this one-step *in vivo* assembly/integration process was around 20% despite the use of small overlaps, of only 65 bp, between the four TUs. The external regions of homology with the genome that allowed to target the final integration of the construct to the rDNA locus were however larger, around 0.6 kb [193]. The screening of successfully engineered yeast cells was facilitated by the orange/red color of positive colonies, due to efficient β-carotene biosynthesis. Some of the best producing transformants, with the deepest color, revealed the presence of an additional integration of some elements from the new pathway, indicating a lack of control of the progress probably imputable to NHEJ events [193]. The same strategy, applied to the assembly of a different 10 kb β-carotene synthesis pathway, in either a control strain and a NHEJ-defective one (*ku70/ku80* double deletion), demonstrated that the assembly/integration was strongly enhanced when NHEJ was inactivated, with an overall efficiency as high as 63% [194]. A similar strategy of one-step *in vivo* assembly/integration was applied at Nanjing Tech University for a 10 kb and three genes arachidonic acid (ARA) metabolic pathway [195]. This work assessed the effect of the overlaps length on the assembly/integration efficiency, demonstrating an optimum of 23% when overlaps of 1 kb were used. The ARA-engineered Po1f strain selected in this study was also shown to exhibit a high growth rate and a genetic stability compatible with industrial use [195]. The HR efficiency of *Y. lipolytica* can thus be sufficient for driving effective *in vivo* assembly and integration of complex structures, especially in a NHEJ-deficient genetic background.

These interesting possibilities have been very recently pushed further in an innovative work from Toulouse Biotechnology Institute (TBI, Toulouse University, France) who have been able to obtain the *in vivo* assembly, in a Po1d derivative, of a 23 kb artificial chromosome [196]. The ylAC plasmid that serves as basis for the whole construct can be digested with different restriction enzymes in order to generate three DNA fragments, each carrying a different selection marker gene, corresponding to the two telomeric ends and to the central region of the future artificial chromosome. These three selectable elements are designed to incorporate, during *in vivo* assembly, two additional DNA fragments corresponding to one or two new metabolic pathways (TUs). Preliminary experiments have been necessary for choosing promoters for the TUs and for determining the optimal ylAC/TUs ratio for *in vivo* assembly. These assays have also established that using an essential gene as additional selection marker was mandatory for the long term stability of the artificial chromosome. An adapted recipient strain, Po1dh, was deleted for 5-aminolevulinate synthase, using CRISPR tools, in order to allow the use of the corresponding *YlHEM1* gene as a selection marker fitted to every cultivation condition [196]. The correct assembly by *in vivo* HR in Po1dh of the *URA3*-bearing left telomeric arm, of the central region bearing the essential *HEM1* gene, of the *LEU2*-bearing right telomeric arm that also carried an ARS/CEN sequence and of two new metabolic pathways was obtained with a more than 90% efficiency. The two TU DNA fragments were obtained by PCR and designed to present 50 bp bordering homologies with the sequences at the extremities of those of the ylAC restriction fragments they needed to be linked to [196]. The proof of concept of the efficiency of the ylAC system was demonstrated by simultaneously engineering *Y. lipolytica* for xylose utilization and for cellobiose catabolism. In less than a week of wet laboratory experiments, the one-step *in vivo* assembly of two heterologous metabolic pathways, each composed of three genes, was performed and the resulting 23 kb artificial chromosome exhibited a stability comparable to that of a natural chromosome [196]. This innovative artificial chromosome strategy represents a powerful new tool for both academic and applied research on *Y. lipolytica* and the ylAC plasmid is intended to

be deposited at Addgene. This remarkable addition to the global *Y. lipolytica* toolbox is expected to be highly beneficial for the engineering of industrial strains.

4.2. Functional elements for the design of expression cassettes

A wide range of functional genetic elements are available in *Y. lipolytica* for the design of TUs adapted to the peculiar purpose of each engineering purpose. They comprise some regulatory components, promoters and terminators, and some optional targeting components that can be fused to the heterologous ORF in order to address the recombinant protein to different organelles or to the vesicular secretion pathway. Secreted proteins can be either released into the cultivation medium or exposed on the surface of the cell (surface display systems). All these functional elements have been extensively described and compared in previous reviews [18, 19, 37, 42] and will only be resumed here. During the early development of *Y. lipolytica* engineering, in the 1980s, the first regulatory and targeting components to be used were those from the few identified genes corresponding to highly secreted enzymes, namely AEP (encoded by the *XPR2* gene) and extracellular LIP2 lipase [17]. The molecular biology progresses now allow to select promising candidate genes by mining genome-wide omics data and to fine-tune the properties of native elements through genetic engineering [42].

4.2.1. Natural *Y. lipolytica* promoters and promoter engineering

The choice of a promoter is of paramount importance for the design of any engineering project: for producing a heterologous protein, strength and either constitutivity or inducibility by a process-friendly inducer are required; for adding a new (or optimizing a native) metabolic pathway, some tunability of the expression for each gene is expected. Among natural promoters, the inducible *XPR2* promoter (noted p*XPR2*), used at first for its high strength, presents a complex regulation [197], not compatible with industrial processes. Other notable natural promoters include the strong constitutive p*TEF* [198] and the inducible p*POX2* [199] and p*EYK1* [200]. Together with its improved version obtained through intron-mediated enhancement, p*TEF*in [201], p*TEF* (from the translation elongation factor 1-◎ gene) still represents a preferred choice for applications that require a strong constitutive promoter. In contrast, known inducible *Y. lipolytica* promoters, mostly derived from lipid pathways (including p*POX2* from an acyl-CoA oxidase gene), have until recently seen their industrial use impaired by incomplete substrate repression and by the hydrophobic nature of their inducers [18, 19]. Fortunately, a study from INRA, Liege University and Gembloux Agro-Bio Tech (Belgium) recently showed that p*EYK1* (from an erythrulose kinase gene) was strongly induced by erythritol and erythrulose, thus providing the non-hydrophobic inducer system that was previously lacking for *Y. lipolytica*. These compounds can be used as non-metabolized inducers in a ΔEYK1 genetic background [200].

The construction of recombinant tailored promoters as an improved alternative to natural ones has been initiated since the 1990s at INRA. Based on a functional dissection of p*XPR2* that highlighted the strength and independency from cultivation conditions of its proximal upstream activating sequence (UAS) [146, 202], a functional DNA sequence, UAS1B, was selected for the design of a series of recombinant promoters [135]. In particular, the hp4d promoter, composed of 4 tandem copies of UAS1B and a core p*LEU2*, demonstrated remarkable properties for heterologous protein production: in culture media of various composition and pH, hp4d demonstrated a high growth-phase dependent expression profile that increases at the beginning of the stationary phase [203]. This peculiar characteristic naturally allows a partial dissociation of growth and expression phases, maximising productivity and minimising potential toxicity concerns. This recombinant hp4d promoter rapidly became, and still remains, the most used worldwide for heterologous expression, as reviewed previously [18, 19, 42]. It has been selected for use in the vectors of the Yeastern YLEX commercial kit (cf. Section 2.3.2).

Thanks to the rapid progress of molecular biology technics, the concept of multi-UASs recombinant promoters has been more recently pushed further at UT Austin: these

authors generalized the concept of "hpNd" tailored promoters by inserting from 1 to 32 copies of UAS1B (aka UAS1$_{XPR2}$) upstream of core promoters from either p*LEU2* or p*TEF*. This new series of recombinant promoters could drive a wide range of expression levels, the strongest ones being as much as eight-fold more powerful than the best natural *Y. lipolytica* promoters [185]. Surprisingly, this study showed that natural yeast promoters could be considered as "enhancer-limited" and that transcription factor availability did not constitute a limiting factor for expression level. Besides UAS1$_{XPR2}$, other enhancer elements have been made available as building bricks for the design of new recombinant promoters, notably the constitutive UAS$_{TEF}$ [204], the erythritol/erythrulose-inducible UAS1$_{EYK1}$ [200] and the fatty acid-inducible UAS$_{POX2}$ [205]. This latter enhancer element was used in the design of a p*POX2*-derived promoter with an unprecedented 48-fold induction level that could be used not only for metabolic engineering but also as a fatty acid biosensor [205]. New strong inducible recombinant promoters were derived from p*EYK1* by tandem addition of multiple UAS1$_{XPR2}$ or UAS1$_{EYK1}$ enhancers [200]. At last, a more recent analysis of p*EYK1* and p*EYD1* (from another erythritol catabolism gene) through phylogenetic footprinting, mutagenesis and hybrid promoter construction using GGA allowed the identification of new enhancers (UAS1$_{EYD1}$ and UAS2$_{EYD1}$) and the design of a series of derived inducible tailored promoters of variable strengths for various applications, from industrial protein production to fine-tuning of gene expression levels in metabolic engineering projects [206]. The tight basal repression level of p*EYD1* allowed the design of derived tailored promoters in which erythritol-based induction could be up to more than 30-fold in a wild-type strain and up to nearly 900-fold in a Δ*EYK1* genetic background [206]. Using the strongest promoter available can however sometime be counterproductive, as was demonstrated in a study comparing recombinant promoters of different strengths, with 2 to 8 UAS1$_{XPR2}$, to produce several secreted proteins of industrial interest: depending on the protein, either the strongest or the weakest promoter produced the best yield, exemplifying how translation and post-translational traffic could constitute limiting steps for some secreted proteins [207]. This study thus suggested that promoter suitability could be protein dependent and that introducing a gene of interest into a pool of vectors carrying different promoters of variable strength can be a powerful strategy to identify the more appropriate one.

Beside the search for new enhancer elements, a consortium of laboratories from Tianjin University (China) proposed an alternative approach for the improvement of recombinant promoters by engineering the core promoter region [208]. The potential role of this region of 20-80 bp between the TATA-box and the transcriptional starting site may have been underestimated, although some early studies had noticed that it may exert some influence on expression [135]. The authors of the Chinese study used a high-throughput approach to screen a library of random 30 bp DNA sequences for their efficiency as artificial core in recombinant promoters directing the bioconversion of lycopene to ⊗-carotene, as reported by changes in color of *Y. lipolytica* colonies. The sequence of these artificial core promoter was shown to strongly influence the strength of the recombinant promoter, with the presence of T-rich elements and a low GC percentage favoring high expression, a feature also found in strong natural yeast promoters [208]. This work allowed a further optimization of recombinant promoters based on p*EXP1* and p*GPD* and added some selected artificial core promoter sequences to the global *Y. lipolytica* toolbox.

4.2.2. Natural and synthetic terminators

Since the 1980s and until recently, terminator sequences obtained from the highly expressed *XPR2* and *LIP2* genes have been used in almost all TU design projects [18, 19]. The role that terminators can play in efficient gene expression has long been underestimated. Being generally considered as necessary but rather neutral elements from TUs, these terminating regions did not raised a lot of interest besides the question of a potential risk of recombination due to their repeated use in multiple TUs. For that reason, the recent INRA Golden Gate toolkit for expression in *Y. lipolytica* includes 13 different terminator biobricks for multiple TUs design [183]. However, the influence of terminator sequences

on transcription completion and mRNA half-life has been rather recently demonstrated in yeasts [209]. As an alternative to natural terminating sequences, some short synthetic terminators designed for *S. cerevisiae*, able to drive a four-fold increase of transcription and expression in this yeast, were shown to be also functional in *Y. lipolytica* [209]. Additionally, the smaller size of these terminators compared to that of natural sequences (35-70 bp versus 100-430 bp) constitutes an advantage for TU and vector design and both their size and synthetic nature minimize the risk of undesired HR between TUs or with the genome. These short synthetic terminators are therefore expected to ease TU construction and to contribute to a higher stability of heavily GM strains.

4.2.3. Targeting the secretion pathway: secretion signals and surface display systems

Vesicular secretion of heterologous or recombinant proteins, which notably facilitates their recovery from the cultivation medium, is obtained by N-terminal fusion of secretion signals to the ORF of the mature protein sequence. In *Y. lipolytica*, the pre or prepro regions from either *XPR2* or *LIP2* genes, encoding the major secreted enzymes, have been used for that purpose since the 1980s, together with *XPR2/LIP2* prepro hybrid sequences [18, 19, 34, 210, 211]. Up to now, the smaller and efficient *XPR2* pre sequence [210] is generally a preferred choice and this secretion signal was selected for the pYLSC1 secretion vector from Yeastern YLEX commercial kit (cf. Section 2.3.2). Alternatively, some heterologous secretion signals have been successfully used in *Y. lipolytica*, especially when from plant or fungal origin, as extensively reviewed previously [18].

More recently, research teams from Poznan University of Life Sciences (Poland) have used genomic and transcriptomic data mining combined to functional screening of a library of GG-built secretion cassettes to identify five new highly efficient secretion signals and to define a consensus sequence for a potentially robust synthetic secretion signal [212]. These new GG-compatible secretion signals constitute valuable additions to the global *Y. lipolytica* molecular toolbox. They were notably used in a synthetic biology study in which a TU and signal peptides shuffling approach allowed the optimization of a starch-utilizing GM *Y. lipolytica* strain for lipid production [189].

The development of surface display methods, for exposing recombinant secreted proteins on the cell surface (arming yeast), is relatively recent in *Y. lipolytica*. Surface display technologies have numerous biomedical and biotechnological applications. Notably, the armed yeasts equipped with various new functionalities can be used as whole-cell microbial factories that can be easily separated from the bioconversion product of interest. The first *Y. lipolytica* surface display method was developed at OUC (China) by C-terminal fusion of the ORF of heterologous genes to a sequence encoding the 110 last amino acids of CWP1, a native cell wall protein [184]. This C-terminal part of *CWP1* gene corresponds to a GPI anchor domain, namely a signal for the posttranslational addition of a GPI (glycosylphosphatidylinositol) structure to the secreted protein and for its covalent linking to β-1,6 glucans from the yeast cell wall (surface display). As reviewed previously [19], a few other GPI anchoring signals from *Y. lipolytica* cell wall proteins have also been applied to surface display in this yeast [213-215], but the OUC/INRA zeta-based auto-cloning pINA1317-CWP110 surface display expression vector [184] remains up to now the more widely used tool for constructing arming *Y. lipolytica* cells.

As reported previously [19], a few GPI-independent alternative strategies have also been applied to surface display on *Y. lipolytica* cells, by fusion of the recombinant protein with the flocculation domain from either *Y. lipolytica* or *S. cerevisiae* Flo1 [213, 216] or with Protein Internal Repeat (Pir) or Chitin Binding Module (CBM) domains from *Y. lipolytica* cell wall proteins [217, 218].

4.2.4. Targeting organelles for subcellular compartment engineering

Instead of being released in the cytoplasm of the cell, recombinant proteins can be specifically addressed to different subcellular compartments through transcriptional fusion of their ORF with the corresponding targeting sequences. They can be retained in the ER if a C-terminal Lys-Asp-Glu-Leu (KDEL) amino acid sequence is present, or targeted

to the cell peroxysomes if a C-terminal tripeptidic peroxisome targeting signal (PTS) is added, the amino acid sequences Ala-Lys-Ile (AKI) or Ser-Lys-Leu (SKL) being the more effective ones in *Y. lipolytica* [45, 83, 219]. Recombinant proteins can also be addressed to LB and oleosomes through transcriptional fusion of their ORF with the C-terminal amphipathic domain of a heterologous oleosin, a structural protein associated with the single layer membrane of plant oil bodies [83, 87]. This oleosome targeting strategy was developed at first for the design of tunable functional nanoparticles that could be assembled in *Y. lipolytica* cells for various biotechnological application [87], as described in Section 2.3.5.

In addition to classical intracellular expression, the possibility of targeting the different recombinant enzymes from a remodelled or newly introduced metabolic pathway to the ER, the peroxysomes or the LB of *Y. lipolytica* cells constitutes a new and powerful tool for fine-tuning the metabolic engineering of this yeast. This novel strategy of compartmentalization of the engineered pathways in different cell organelles has been fully exploited in a recent work from the Huazhong University of Science and Technology (Wuhan, China): the simultaneous targeting of an engineered lipase-dependent metabolic pathway to ER, peroxisomes and LB allowed to maximize the yield of biofuels produced in the resulting GM *Y. lipolytica* strains, as evoked in Section 2.3.4 [84]. Such recent possibilities of compartmentalized metabolic engineering will facilitate complex engineering strategies that aim at maximizing the yield of a compound of interest by a holistic approach, taking into account not only the targeted pathway and its possible bottlenecks but also the necessary up-regulation of the cofactor and/or anti-oxidative pathways

4.2.5. Selection marker genes and marker rescue systems

Various marker genes can be used for GM strain selection in *Y. lipolytica*: several genes from amino acids or uracil biosynthesis pathways for auxotrophy complementation, some resistance markers and a few catabolic markers that can be selected by the ability to use a given substrate as sole carbon or nitrogen source [18, 19, 42]. If auxotrophy markers are the most commonly employed due to their historical incumbency and their simplicity of use, they require the preliminary construction of the corresponding auxotrophic strains, which constitutes a drawback for the valorisation of natural strain biodiversity. In contrast, catabolic and resistance marker genes constitute dominant selection markers that can more easily be applied on a larger scale to the screening of strains for a given application.

Among the available auxotrophy markers, *LEU2* and *URA3* genes have been and remain the most commonly used, according notably to the popularity of the Po1 series of recipient strains bearing *leu2* and/or *ura3* non-leaky non-reverting mutations (cf. Section 3.1.3 and Figure 2). Other less common choices are *LYS5* and *TRP1* genes, as reviewed recently [42]. The availability of *Y. lipolytica* genomic sequences and the new possibilities for easier engineering, such as CRISPR tools (cf. Section 4.3), are expected to facilitate a future increase of the range of auxotrophic markers and corresponding multi-auxotrophic strains. Despite the popularity of *LEU2* as a selection marker, some rather recent studies have raised concern about a possible impact of leucine metabolism on lipogenesis in *Y. lipolytica* [153, 220], a factor that pleads in favour of the use of other auxotrophy or preferably dominant markers, at least for lipid-related metabolic engineering projects.

The *URA3* gene constitutes since the 1990s a favorite choice when a marker rescue step is to be used, thanks to its easy counter-selection using the toxic uracil precursor analog 5-FOA (5-Fluoroorotic acid) [221]. Marker rescue consists in restoring an auxotrophy by deleting an integrated selection marker, for allowing its repeated use in multiple sequential engineering steps. More recently, a counter-selection system has also been made available for the *TRP1* gene in *Y. lipolytica*, using the toxic analog 5-fluoroanthranilic acid (5-FAA) [222]. The Cre-lox recombination system from bacteriophage P1 has been used at INRA to design a user-friendly system for gene deletion and marker rescue in *Y. lipolytica*: any integrated loxP-bordered marker TU can be efficiently excised through transient expression of the heterologous Cre recombinase using a replicative vector [178]. Different loxP-excisable markers have been included in some of the newly developed *Y. lipolytica*

toolboxes, the two EasyCloneYALI toolkits [190] and the Golden Gate toolkit CRISPR tools [223]. Alternatively, marker rescue can be obtained through a blaster cassette strategy, by making use of the enhanced HR rate of a $\Delta Ku70$ genetic background to allow the use of short repeated DNA bordering sequences for excising a marker for which a counter-selection is available. Notably, a TRP-blaster system and a URA3-blaster cassette using 100 bp bordering repeats have been described [222, 224].

Unexpectedly, some transposomics tools, more generally associated with massive genome studies like Tn-seq (cf. Section 4.3.4), can also be used to enable scarless precise excision of a selection marker following integration of a TUs-bearing cassette by classical or CRISPR-based methods. Conceived as an alternative application of the *in vivo* transposition system based on *piggyBac* transposon described below (cf. Section 4.3.4), this tool can constitute an efficient marker rescue system following classical or CRISPR-mediated marker-selected integration of exogenous pathways, by allowing scarless removal of the selection marker, for reiterated use or safety regulation purposes. This alternative use of the *piggyBac in vivo* transposition system makes use of an excision-only transposase variant, heterologously expressed from a replicative vector, to obtain the excision from the genome of any DNA fragment bordered by *piggyBac* inverted terminal repeats (ITRs) [150].

Although naturally resistant to most common antibiotics, *Y. lipolytica* cells are sensitive to hygromycin B, to nourseothricin, to glycopeptide antibiotics of the bleomycin/phleomycin group and to mycophenolic acid, so that the corresponding *hph* (Hyg[R]), *NAT* (NTC[R]), *ble* (Phleo[R]) and *guaB* genes can be used as dominant markers [150, 190, 225, 226]. The Hyg[R] gene indeed constitutes a preferred marker in the INRA gene disruption and marker rescue Cre-lox-based system [178]. Together with NTC[R], this dominant marker was included in the two EasyCloneYALI toolkits [190] and in the Golden Gate global toolkit [183, 223].

Interestingly, the *ScSUC2* TU integrated in W29-*ura3-302* and the derived Po1 series of strains (cf. Section 3.1.3 and Figure 2) was intended at first to be used as a catabolic dominant marker when sucrose was used as sole carbon source [14], but was preferably applied to recipient strain improvement through sucrose consumption [9, 42]. An optimized version of this invertase TU using the p*TEF* and its native secretion signal [227] was included as dominant marker in the Golden Gate toolkit [183]. A second catabolic dominant marker, using a *DsdA* TU encoding a bacterial D-serine deaminase to allow cell growth on D-serine as sole nitrogen source [228], has been included into the Easy-CloneYALI toolbox [190].

These dominant resistance or catabolic markers can be used to engineer directly wild-type strains, which could allow leveraging *Y. lipolytica* natural biodiversity for biotechnological application. Such tools could enable selecting among a large range of wild-type isolates for peculiar characteristics well-adapted to each intended application, an objective pursued by more and more research teams [100, 144, 223, 229]. They could also be seen as potential tools to be assayed for attempting to engineer other yeasts from the *Yarrowia* clade for biotechnological applications, as evoked in Section 2.4.2 [99, 100, 103].

4.3. Gene editing and whole genome analysis technologies: CRISPR and other tools

Since not more than a decade, a number of innovative genome editing and massive analysis strategies have emerged, based on different bacterial defense systems, notably on the very popular 2020 Nobel-Prize-winning CRISPR system, or on some *in vivo* transposition systems. These new tools have rapidly been adapted for use in numerous organisms, including *Y. lipolytica*, as reviewed previously [42, 230] and resumed below.

4.3.1. CRISPR tools and *Y. lipolytica* CRISPR-Cas9 toolboxes for gene editing

The numerous CRISPR (clustered regularly interspaced short palindromic repeats) tools that have been derived from the bacterial Cascade (CRISPR-associated complex for antiviral defense) system are all based on the production, into the cell nucleus, of a nuclease (generally Cas9) complexed with a synthetic single guide RNA (sgRNA) that targets

it to the homologous genomic locus into which a double-strand break (DSB) needs to be introduced [231]. This DSB will be repaired either by NHEJ, generating insertion/deletion (indel) mutations or, if a donor sequence is present, by homology-directed repair. In presence of a TU(s)-bearing DNA fragment flanked by homologies to the genomic DSB site, this cassette will serve as a repair template so that the TU(s) will be integrated by HR at the targeted locus. Such a powerful CRISPR–Cas9 tool can be used for example to obtain, without the need of a selection marker, the disruption of a target gene and the simultaneous integration of a new metabolic pathway [231]. Adapting CRISPR tools to a new organism can constitutes a challenge, since it necessitates the fine-tuned expression of both Cas9 nuclease and sgRNAs, together with the intranuclear targeting of the heterologous enzyme under a fully functional form. The first *Y. lipolytica* CRISPR- Cas9 systems have been developed independently in 2016 during a collaboration between California, Riverside and Clemson Universities (USA) and by a consortium of Shanghai laboratories [232, 233].

The American research teams combined on a unique pCRISPRyl replicative vector the expression of a codon bias optimized *Cas9* gene from an 8UAS1-pTEF promoter and the expression of the sgRNA of choice from a recombinant promoter, fusion of a native RNA-PolIII promoter with a tRNA, allowing to use endogenous tRNA processing for improved mature sgRNA production. This pCRISPRyl vector, which can be applied to CRISPR-Cas9 markerless gene disruption and/or targeted integration of a HR donor sequence, was deposited at Addgene [232]. The same authors also screened *Y. lipolytica* genome for loci that could integrate heterologous TU with no impact on cell growth and five of the corresponding sequences were included in the CRISPR toolkit, as five pairs of vectors targeting each of these neutral loci (for each selected locus, a pCRISPRyl vector expressing the corresponding sgRNA and a corresponding homology donor vector, carrying a GFP TU as fluorescent tag, all available from Addgene). The five selected loci include the two extracellular protease *XPR2* and *AXP1* genes, which disruption is beneficial for heterologous protein production (cf. Section 3.1.3). A proof of concept of these CRISPR tools was made through rapid integration of four genes from a semisynthetic metabolic pathway for lycopene production, at four of the neutral loci [234]. The co-transformation of pCRISPRyl and of a donor vector demonstrated markerless HR integration of the donor cassette with a 64% efficiency in a Po1f strain and a 100% one in a NHEJ-disrupted Δ*Ku70* derivative [232]. The Shanghai research teams also designed a unique replicative vector to express *Cas9* and the desired sgRNA, both from p*TEF*in promoters (pCAS1yl vector, deposited at Addgene), with the possible addition of a homologous donor DNA cassette on the same plasmid (pCAS2yl vector) [233]. The pCAS1yl vector demonstrated markerless gene disruption (indel knocking-out by NHEJ) with a more than 85% efficiency in a Po1f strain. Simultaneous double or even triple gene disruptions were also demonstrated, using a pCAS1yl vector carrying two or three tandem sgRNA TUs [233].

More recently, the Novo Nordisk Foundation Center for Biosustainability (Denmark) added some CRISPR tools to its multipurpose EasyCloneYALI toolbox (cf. Section 4.1.3): the EasyCloneYALI toolkit for gene editing comprises a series of 15 pCfB vectors for gene deletion and/or markerless TU integration, all available from Addgene [190]. This Easy-CloneYALI CRISPR toolkit, includes two integrative vectors for *Cas9* expression from p*TEF*, carrying a loxP-excisable Hyg[R] or DsdA dominant marker, a replicative vector for expression of the sgRNA of choice from p*Pot1*, carrying a loxP-excisable NTC[R] dominant marker, six disruption vectors with sgRNAs targeting six selected loci and the six corresponding donor vectors for integration of the TU(s) of choice at these loci. As a proof of concept of the stability of the GM strains obtained with this CRISPR system, as much as eleven steps of successive TU integrations were performed with no detectable loss of the previously integrated TUs [190].

A consortium of French research teams from INRA, INSA Toulouse and Toulouse University also designed a *Y. lipolytica* CRISPR-Cas9 system and applied a holistic approach to studying the NHEJ-based gene knocking-out process, using multiparameter flow cytometry combined to genotypic and phenotypic analyses [235]. These authors

demonstrated that the limiting factor during the RNA/protein complex formation was not Cas9 nuclease availability but sgRNA design, sequence and level of expression, while providing a more general insight into the metabolism of small RNAs in yeast cells. The genome of a *Y. lipolytica* knocked-out strain was then fully sequenced, with no evidence of undesired sequence change, which confirmed the safety of CRISPR-Cas9 tools for gene editing in *Y. lipolytica* [235].

Some CRISPR-Cas9 tools compatible with the GGA method (cf. Section 4.1.3) have also been designed recently for *Y. lipolytica*. A consortium of Austrian laboratories developed a GoldenMOCS-Yali toolkit for expression and CRISPR-Cas9-based metabolic engineering in *Y. lipolytica*, explicitly aimed at leveraging natural biodiversity [229]. Their vectors are compatible with previous GoldenMOCS tools, corresponding to a rapid GG cloning strategy used in multiple species. A CRISPR-Cas9 GoldenMOCS-Yali vector was used for knocking-out a gene into different *Y. lipolytica* wild-type strains, with an efficiency ranging from 6% to 25%, depending of the genetic background [229]. Similarly, the stated intention of valorizing the biodiversity of wild-type *Y. lipolytica* strains prompted the design at INRAE of a series of five GGA-compatible CRISPR-Cas9 vectors, derived from the pCRISPRyl vector, with different loxP-excisable selection markers that include the dominant Hyg[R] and NTC[R] [223]. These CRISPR tools are notably compatible with the pool of GG bricks described in Section 4.1.3 [183]. As a proof of concept of the possibility for wild-type strain genome editing, the knocking-out a gene was assayed in nine wild-type isolates and was achieved in seven of them, with a high efficiency in four strains and a 100% success rate in two [223]. These two new sets of CRISPR tools, both available from Addgene, are expected to allow leveraging *Y. lipolytica* natural biodiversity by allowing to select the more appropriate wild-type isolate for each intended application and then to engineer rapidly and easily its metabolism.

However, the low rate of HR in *Y. lipolytica* limits the capacity of this yeast for homology-mediated DNA repair during site-specific CRISPR gene editing. This drawback has been addressed by a MIT research team who improved the CRISPR/Cas9-based methods using tRNA processing for sgRNA production (pCRISPRyl and derivatives) by redesigning the tRNA-sgRNA fusion system through the use of secondary RNA structure prediction. This strategy for improved sgRNA expression resulted in a higher efficiency of CRISPR/cas9 gene editing at chromosomal loci for which it failed or was ineffective when using previous methods [236].

Besides Cas9, Cpf1 (aka Cas12a) is another nuclease that can be applied to designing CRISPR tools. In contrast to Cas9, which uses naturally two short CRISPR RNA instead of an artificially designed sgRNA, Cpf1 is naturally a single RNA-guided endonuclease. Cpf1 recognizes a less frequent T-rich PAM (protospacer-adjacent motif), instead of a G-rich one for Cas9, and generates sticky ends DSB more distal to the recognition site. These characteristics are expected to allow a more reliably targeted, and possibly repeated, genome editing, which prompted the design of a CRISPR-Cpf1 system for *Y. lipolytica* editing at the University of Maryland, Baltimore County (USA) [237]. This CRISPR-Cpf1 system allowed to perform indel NHEJ-based gene knocking-out with an efficiency in the range of 95% and multiplex editing with a mean efficiency of 79% for double targets and 42% for triple ones [237].

Whatever the CRISPR tool used, the choice of sgRNA sequences remains a crucial step for successful gene editing (or transcriptional control). A study of the relative efficiencies of six candidate sgRNA sequences for each *Y. lipolytica* gene has recently been performed by a consortium of American laboratories, who constructed a library of sgRNAs targeting the 300 first nucleotides of 7854 ORFs from Po1f strain and measured their individual DSB-promoting efficiency in a CRISPR-Cas9 editing assay [238]. This study revealed that 48% of the sgRNA from the six-fold coverage library exhibited a high efficiency and that 95% of *Y. lipolytica* ORFs could be targeted by at least one very highly efficient sgRNA [238]. As reported in more details in Section 4.3.4, this work also provided an innovative multipurpose whole genome analysis method for *Y. lipolytica*.

4.3.2. TALEN tools for gene editing in *Y. lipolytica*

Besides CRISPR tools, some alternative genome editing strategies are available and a few have been adapted for use in *Y. lipolytica*. TALEN (transcription activator-like effector nucleases) are recombinant restriction enzymes designed by fusing a nuclease with a TAL effector DNA-binding domain that can be engineered to target specific DNA sequences [239]. TALEN are used to generate, at specific genomic loci, DSB that can be repaired through NHEJ, or HR in presence of a homologous donor sequence, like for CRISPR tools. However, for TALEN, the DNA recognition system is part of the recombinant enzyme when, for CRISPR tools, it is brought by the sgRNA. Therefore, the design of a new TALEN for each targeted locus is laborious and costly compared to the more rapid and cheap synthesis of a small RNA. Despite this drawback, a TALEN tool has been used at INSA (Toulouse University) for structure-based mutagenesis of the ketoacyl synthase domain from the *Y. lipolytica* multifunctional fatty acid synthase (FAS), a key enzyme of lipid synthesis pathways, to allow the production of fatty acids with shorter chain lengths [240]. A step-by-step methodological guide of how-to-use TALEN in *Y. lipolytica* for knocking-out or introduction of point mutations into genes was very recently made available by the same authors [241].

4.3.3. Other CRISPR tools for base editing and for gene repression or activation

The targeting ability of the CRISPR-Cas9 system has been exploited in different derived methods in order to allow recruiting new functionalities at a precise genomic locus. The construction of a recombinant Cas9 nuclease fused with an activation-induced cytidine deaminase (Target-AID) has been applied at first for targeting point mutagenesis at precise loci from plant genomes [242]. This innovative method for targeted base editing has been adapted recently to *Y. lipolytica*, at the Seoul National University (South Korea) and applied to multiple gene disruption [243]. The mechanism of this *Y. lipolytica* Target-AID base editor is to use the Cas9/sgRNA complex to recruit a cytidine deaminase (CDA) able to provoke a C to T mutation at the chosen locus, with the aim to create a stop codon within the targeted ORF. In order to obtain (multiplex) gene disruption in a Po1g *ku70Δ::loxP* strain, the sgRNA(s) and a recombinant protein, corresponding to Cas9 fused to both a heterologous CDA and a uracil DNA glycosylase inhibitor, were simultaneously expressed from a pCRISPRyl-derived vector [243]. Using a Δ*Ku70* genetic background is required in order to increase the accuracy of the mutagenesis process by preventing the formation of indels following the cytidine deamination step. The optimization of the fusion enzyme expression level of this Target-AID system enabled performing single gene disruptions with a 94% efficiency and double gene ones with a 31% efficiency [243].

Several CRISPR tools for gene repression or activation have also been designed, which make use of defective versions of Cas9 (dCas9) or Cpf1 (dCpf1) lacking endonuclease activity while retaining their sgRNA-directed DNA targeting function. The University of California, Riverside (UCR) applied their *Y. lipolytica* CRISPR tools (pCRISPRyl replicative vector, cf. Section 4.3.1) to gene repression through the CRISPR interference (CRISPRi) system [244]. At first, they simply used dCas9 to sterically repress gene transcription from sgRNA-targeted loci, with a proof of concept demonstrated by obtaining NHEJ-deficient *Y. lipolytica* strains by directing dCas9 to both *Ku70* and *Ku80* promoters, using multiplex sgRNAs [244]. They were then able to improve the efficiency of gene repression by using a fusion protein corresponding to dCas9 linked to the transcription repressor Mxi1 (Schwartz et al., 2017b). The corresponding CRISPRi vectors (pCRISPRi_Mxi1_yl), empty or bearing the *Ku70* and *Ku80*-directed sgRNAs for NHEJ inactivation, are available from Addgene [244]. A similar CRISPRi system for gene repression in *Y. lipolytica* was developed at Tianjin University, which uses either dCas9 or dCpf1, each alone or fused with a KRAB repressor [245]. This new CRISPRi system uses a one-step GGA strategy and takes advantage of multiplex sgRNA targeting to repress simultaneously the transcription of multiple genes, or to repress more effectively that from a single gene by aiming at several promoter sites. Namely, a high level of repression was achieved (85% when using

dCpf1 and 92% with dCas9) when using three sgRNAs to target different sites in the promoter of an integrated *gfp* reporter gene, a strategy that allows avoiding the preliminary screening step usually necessary to select the more effective sgRNA sequence [245]. These CRISPRi tools constitute very valuable additions to the global CRISPR toolbox, as efficient ways to transiently repress any *Y. lipolytica* gene activity, and will find a preferred application for temporary repression of NHEJ activity prior to some major genetic engineering step requiring a high HR rate. They will in that case allow avoiding a permanent deletion of *ku70/ku80* genes that could impair the robustness of the resulting GM strain.

At last, the CRISPR toolbox from UCR has also been completed with some CRISPR activation (CRISPRa) tools, in a collaborative work with UT Austin. Like for CRISPRi, CRISPRa is based on the fusion of a defective nuclease (dCas9) with a transcriptional effector (activator) and makes used of a sgRNA to direct the new (transcription enhancing) functionality to the targeted gene. The synthetic tripartite activator VPR was selected among several candidate activators and, when fused to dCas9, enabled to enhance the transcription of two dormant native β-glucosidase genes, conferring to the GM strain the ability to growth on cellobiose as sole carbon source [246]. The corresponding pCRISPRa_VPR_yl vector is available from Addgene. This new CRISPRa tool pushes further the versatility of the CRISPR-Cas9 system by providing an easy and elegant way to investigate the potential of transcriptionally silent parts from the *Y. lipolytica* genome. As already demonstrated in previous studies for the pentose pathway [55, 56] (cf. Section 2.3.3) and confirmed here for cellobiose, mining the *Y. lipolytica* genome for dormant metabolic pathways is a promising strategy for rewiring its metabolism for alternative substrates, and the CRISPRa tools could greatly facilitate this approach. The combined use of CRISPRi and CRISPRa systems is expected to enable fine-tuning *Y. lipolytica* metabolic pathways and to ease complex engineering project. A methodological guide of how-to-use these new tools was very recently made available [247].

More generally, UT Austin also proposed a methodological guide for using the whole series of CRISPR-Cas9 vectors they developed (or contributed to develop) for either genome editing or dCas9-based transcriptional control [248].

4.3.4. Transposomics and CRISPR-derived tools for whole genome analysis in *Y. lipolytica*

The development of whole genome analysis tools applicable to functional genomics studies is rather recent in *Y. lipolytica*, despite their interest for holistic approaches of metabolic engineering strategies. The basic idea of such tools is to obtain a genome-wide library of viable mutant cells through a tagging method that allows subsequent massive sequencing of the junction sequences and identification of the corresponding mutated genes. The comparison of the results obtained following growth of the mutant library in various conditions allows to obtain an "inverted image" of the genes contributing to a function of interest. Namely, the surviving tagged mutants can be identified as non-essential and having therefore a low fitness coefficient for the tested growth condition. A preferred whole genome analysis method is transposon insertion sequencing (Tn-seq), which apply transposomics tools to obtaining genome-wide transposon-generated insertional mutants libraries, and generates fitness coefficients through massive sequencing of transposon insertion junctions.

The first transposomics tools applicable to whole genome analysis have been developed only in 2018 for *Y. lipolytica*, based on a *piggyBac* transposon at UT Austin [150] and on a Hermes transposon (from the dipter *Musca domestica*) at the University of California, Irvine (UCI, USA) [249]. Such tools are based on the mobilization of a modified transposon, corresponding to a selection marker bordered by the transposon ITRs, using a heterologously expressed dedicated transposase (generally adapted to *Y. lipolytica* codon bias and born on a replicative vector). As the *piggyBac* transposition system targets specifically TTAA sequences that are present in less than 2/3 of *Y. lipolytica* ORFs and is also biased in favor of transcribed regions, it presents only a limited interest for whole genome saturation mutagenesis applications such as Tn-seq [150]. This transposomics tool is consequently preferably proposed as a cargo-mobilizing system for marker rescue, as described

above in Section 4.2.5. For developing Tn-seq applications, a transposon that integrates as randomly as possible into genomes is required in order to maximize genome coverage. However, the final success of a Tn-seq system will depend on the effective expression of a functional heterologous transposase, a step that can be challenging.

The UCI team have used the Hermes transposon for performing genome-wide saturation mutagenesis into W29 strain. They constructed a library of more than 500,000 insertion mutants that was cultivated for 80 generations on either glucose or glycerol as sole carbon source, before being submitted to massive sequencing in order to identify the mutants present in the surviving population. A statistical analysis of the resulting data identified 22% of *Y. lipolytica* genes as essential and allowed to measure the contributions of non-essential genes to growth on each carbon source [249]. These data were used to evaluate and sometimes correct two previously established genome-scale models of *Y. lipolytica* metabolism [250, 251]. A selection of the insertion mutants in which lipid accumulation was enhanced was performed by combining Bodipy lipid staining and fluorescence-activated cell sorting (FACS). The identification of the corresponding genes provided new insights into *Y. lipolytica* lipid metabolism and new targets for future metabolic engineering [249]. The vectors for *in vivo* Hermes transposition used in this Tn-seq study are available from Addgene. They include a replicative vector carrying both the Hermes transposase TU and the modified transposon (ITR-bordered *LEU2* marker), a markerless version (for inclusion of any marker TU) and some negative control vectors. Depending on the growth conditions compared, this Tn-seq system can be applied to various whole genome studies of *Y. lipolytica* strains for dissection of their metabolism or evaluation of their robustness for biotechnological applications.

Besides transposomics tools, other methodologies can be applied to whole genome analysis, as long as they allow building a large library of tagged mutant strains. Somewhat surprisingly, the precisely targeted CRISPR tools for gene inactivation can be used for such studies, providing the preliminary construction of a sgRNA library covering the whole genome. This strategy has been chosen by a consortium of American laboratories who were also able to quantify the cutting efficiency of each sgRNA and thus to identify the best sgRNA for each *Y. lipolytica* gene, as described in Section 4.3.1 [238]. The six-fold coverage sgRNA library constructed by these authors was used to transform in parallel three strains with the same Po1f background: a control and two *Cas9*-expressing derivatives, with or without NHEJ disruption (Δ*Ku70*). In presence of a functional NHEJ system, the DSB generated could lead to indel formation, allowing to construct a library of viable mutant cells and to determine a fitness coefficient for each gene. In contrast, the DSB cannot be repaired in the Δ*Ku70* strain, no donor HR sequence being present, which leads to cell death and thus to the obtention of a cutting score for each sgRNA sequence [238]. By comparison of the progeny obtained in presence or absence of native DNA repair, this work provided a high-throughput measurement of each sgRNA-directed cutting efficiency and allowed identifying highly efficient sgRNA sequences for nearly 95% of *Y. lipolytica* genes (cf. Section 4.3.1). This study also improved the identification of *Y. lipolytica* essential genes by a significant reduction of false negatives thanks to the sgRNA redundancy. In addition, the library of viable mutants obtained in *Cas9*-expressing Po1f strain can be used to determine the fitness score of each gene under various cultivation conditions. This CRISPR library was notably applied to the study of lipid metabolism, through the use of a fluorescent lipid dye and of a FACS high-throughput screening system, and allowed identifying new target genes which disruption favors lipid accumulation [238].

4.4. How gene editing can leverage strain biodiversity and be a source of new engineering strategies

A few decades ago, laboratories working on *Y. lipolytica* engineering tended to focus on a selected strain and to invest a lot of work and time into introducing one, two or three auxotrophies in its genome in order to be able to use the corresponding genes as markers for further engineering. The first step was generally to obtain an Ura- derivative, thanks to the easy counter-selection of uracil-deficient strains with 5-FOA. The more recent alter-

native use of dominant markers remained limited and the implementation of marker rescue systems was required for complex engineering projects. In contrast to the more recently developed scarless methods (*piggyBac*, marker-blasters) [150, 222, 224], these former marker rescue technics presented the drawback of leaving some genetic manipulation vestiges in the genome. Moreover, nowadays, the availability of markerless CRISPR-based genome editing tools [190, 223, 229, 232, 233] makes possible the rapid engineering of numerous strains for testing a desired application. As discussed in Section 4.3.1, these new tools can make the valorization of *Y. lipolytica* biodiversity an attainable goal and some of these CRISPR-Cas9 tools, compatible with rapid assembly methods, have been explicitly designed for this purpose [223, 229]. These approaches could benefit from the large range of wild-type *Y. lipolytica* isolates made available for research purposes by the yeast collections listed in Table 2. The promising demonstration that some of the tools developed for *Y. lipolytica* could be successfully used to engineer *Y. phangngensis* [104], is also a good omen for the future possibility of extending their usage to other yeasts of the *Yarrowia* clade. Besides this direct impact of CRISPR-based and other recently developed tools for markerless genome editing, they can also be a source of new engineering strategies by allowing or easing some complex genomic changes, notably for NHEJ disruption or for mating type switching.

4.4.1. Increasing the homologous recombination efficiency in *Y. lipolytica*

As already evoked several times above in this review, *Y. lipolytica* uses predominantly its NHEJ system for the repair of DSB, so that the flanking regions of homology with the genome of a DNA cassette need to be large enough (0.5–1 kb) to allow a correctly targeted integration [9, 178]. We have also seen several examples of how this drawback could be palliated by knocking-out *Ku70* and/or *Ku80* gene(s) [194, 222, 224, 232, 238, 243]. This strategy dates from 2013, when these gene deletions (performed using classical molecular biology methods) have been used to allow facilitating ulterior genetic engineering of Po1d and H222 strains [252, 253]. A series of Δ*Ku70* strains derived from Po1g, with different auxotrophies, have also been included in a *Y. lipolytica* toolbox [149]. However, all these NHEJ-disrupted strains exhibit a reduced transformation efficiency and the permanent knocking-out reduces their robustness, as reviewed previously [43].

Some CRISPR-derived tools can now offer an alternative strategy of transient NHEJ repression that allow to combine the ease of genetic engineering with the possible reversion to a more robust NHEJ-prone phenotype when needed. Namely, the CRISPRi tools described in Section 4.3.3 could find a preferred application in the transient repression of *Ku70* and/or *Ku80* genes, in order to reduce *Y. lipolytica* NHEJ activity prior to genetic engineering, without the drawbacks linked to a permanent knocking-out [244, 245].

At last, an alternative approach for increasing homologous recombination in *Y. lipolytica* has been recently proposed by a consortium of Chinese and British laboratories: these authors expressed in Po1f strain a codon-adapted *RAD52* gene from *S. cerevisiae*, encoding the main recombinase from this HR-prone yeast [254]. The heterologously expressed *ScRAD52* improved HR efficiency in *Y. lipolytica* Po1f strain more efficiently than a Δ*ku70* knock-out. The effect was maximized in a Po1f-Δ*ku70-ScRAD52* strain, in which the HR efficiency reached 95% when using homology arm lengths of 1 kb [254]. It would be interesting to combine *ScRAD52* heterologous expression with *Ku70/Ku80* CRISPRi-mediated transient inactivation to possibly keep the best of these HR enhancement strategies.

4.4.2. Diploid strain formation and sexual hybridization following mating type switching

As evoked above in Sections 2.2.2 and 3.1.2, the frequencies of mating between Mat-compatible wild-type isolates and of hybrid sporulation are very low, probably due to genomic differences linked to the presence of retrotransposons and other retro-elements. The formation of diploid strains would however present several advantages for the design of cell factories, since they exhibit a more robust growth than haploid ones and they can combine the interesting properties from different GM parent strains. Therefore, some

early mating-type switching experiments had been reported, as a mean to obtain diploid strains by deleting *MATA* and inserting *MATB* through classical molecular biology methods, in order to obtain the sexual hybridization of a mating-type switched strain with its parental type [255]. If these first assays demonstrated the possibility of obtaining a functional "trans-sexual" strain, the observed mating frequency remained however very low and there was no other report of mating-type switching in *Y. lipolytica* during the following decade. Thanks to the development of new scarless gene editing technologies, this problem of mating type switching has recently been overcome, independently, by a research team at UT Austin and by a consortium of South Korean laboratories, who obtained the efficient formation of diploids between GM strains issued from a same genetic background [58, 256].

As evoked above in Section 2.3.3, the UT Austin research team constructed xylose-utilizing strains (XUS) from PO1f and E26 (an obese derivative) strains. They were able to switch the mating type of the E26-XUS strain from A to B, by scarless site-specific HR at the mating type locus using the *piggyBac* cargo-mobilizing system for marker rescue (cf. Section 4.2.5). The crossing of this mating-type switched strain with each of three E26-based type A production strains, genetically engineered to produce respectively α-linolenic acid (ALA), riboflavin or triacetic acid lactone (TAL), allowed generating diploid strains producing these new metabolites from xylose as sole carbon source, with yields similar or higher than those of each parental strain from glucose [58]. This sexual hybridization approach could thus allow an easy merging of the properties of two GM strains, which could highly facilitate complex metabolic pathway engineering. The diploid production strain obtained also showed a robust growth and a high stability, demonstrating the high potential of this strategy for the development of efficient cell factories. The satisfactory level of mating frequency observed in these experiments plead in favour of a former detrimentous effect of vestiges that classical genetic manipulations could leave in the genome, a factor that was this time avoided through the use of a more recent scarless gene editing method.

Similarly, the South Korean research teams obtained a mating type switch from A to B in Y2–1U, a GM strain derived from Po1g by deletion of *ku70* and of the six *ACO1-6* β-oxidation genes [256]. The mating type switching was obtained by scarless site-specific HR at the mating type locus using an *URA3*-blaster marker rescue system in a *ΔKu70* genetic background. The self-mating of this GM strain devoted to the production of dodecanedioic acid, a monomer for polyester and polyamide, was aimed at augmenting its ploidy for obtaining a more robust and efficient cell factory [256]. This second example confirmed that the innovative strategy of diploid formation through mating type switching could ease pathway engineering and bioprocess development. Increasing strain ploidy is indeed expected to enhance genetic stability, stress resistance and productivity of cell factories [256].

4.5. Towards a holistic view of cell factories and bioprocesses development

4.5.1. The *Y. lipolytica* pan-genome

To the best of our knowledge, the sequences of a total of 18 *Y. lipolytica* strains have been published up to now (some others being not publicly accessible), namely, chronologically, those of E150, Po1f, W29, WSH-Z06 and five KGA hyper-producer mutants, A-101, a *ku70* mutant of Po1g, YB-392, YB-419, YB-420, YB-566, YB-567, H222 and YlCW001 (ionic liquid-resistant laboratory-evolved derivative of Po1f) [25, 26, 28, 29, 107, 132, 136, 137, 141, 144, 149]. As some of these strains share a common genetic background (W29 with its derivatives Po1f and g and their mutants; WSH-Z06 with its mutants), this represents only ten independent genomes, but can however constitute the start of a pan-genome, representative of *Y. lipolytica* genetic diversity. Besides usual bio-informatics tools for genomics analysis, the study of eukaryotic microbial pan-genomes requires specific data pipelines, such as the Pangloss tool recently developed at Maynooth University (Co. Kildare, Ireland). This software pipeline can be used to construct a pan-genome from a set

of genomes from any eukaryotic microbial species, using a PanOCT (Pan-genome Ortholog Clustering Tool) approach, and be applied to various functional characterisation and visualisation analyses of this pan-genomic data. These authors used Pangloss to design a *Y. lipolytica* pan-genome based on a set of seven sequences selected from wild-type isolates of various origins [257]. The *Y. lipolytica* pan-genome still contains a large majority of ORFs which function remains unknown or only putative (http://gryc.inra.fr/). The recent development in *Y. lipolytica* of genome-wide high-throughput analysis technics, such as Tn-seq or CRISPR-based ones (CRISPR-directed knocking-out, interference or activation) is expected to allow reducing this number in a near future. As shown above in Sections 2.3.3, 3.1.4 and 4.3.3 for respectively pentose, galactose and cellobiose [55, 56, 156, 246], mining the *Y. lipolytica* pan-genome for dormant (or inefficient) metabolic pathways could also allow to improve the consumption of alternative substrates, notably through the use of CRISPRa tools.

4.5.2. Genome-scale omics tools and metabolic models

As several examples have been given throughout this review and as is tentatively schematized above in Figure 3, new tools and new strategies for metabolic engineering can now allow to envision more ambitious metabolic engineering projects, by implementing efficient *in vitro* and *in vivo* assembly methods as well as a wide range of gene editing and regulating tools, notably the CRISPR-derived ones (cf. Sections 4.1.3, 4.1.4 and 4.3). Metabolic engineering projects are now expected to take into account the whole cellular metabolism, including the intracellular metabolic fluxes, the availability of cofactors and the effect of oxidative compounds, and to use the possibility to compartmentalize the metabolic modifications in the different cellular organelles (cf. Section 4.2.4). Nevertheless, any cell factory design rely at first on a precise understanding of cellular metabolism that can only be built through genomic, transcriptomic, metabolomic or fluxomic analyses [258-261]. White biotechnologies have thus benefited from the rise of multi-omics technologies that allowed obtaining *in silico* models of genome-scale metabolic pathways, giving a more holistic point of view on cell factory design. Bioinformatics and applied mathematics have thus their role to play, through the design of GEMs (genome-scale metabolic models) in the deciphering of *Y. lipolytica* regulatory networks [250, 251, 262-265]. Some more recently developed genome-wide high-throughput tools, such as the Tn-seq and CRISPR-based approaches described in Section 4.3.4 will increase our understanding of cellular processes and allow to verify the predictions of the GEMs [238, 249], a first step towards an improved second generation of such models. In one of the most recent of these GEM studies, a consortium of French laboratories, combined a network interrogation process with a validation step through bench experimentation in order to identify regulatory elements and mechanisms promoting lipid accumulation in *Y. lipolytica* [264]. This study allowed the design of YL-GRN-1, a gene regulatory network comprising more than hundred transcription factors, four thousand target genes and seventeen thousand regulatory links. Nine new potential regulators of lipid accumulation were identified, from which six were validated in subsequent wet laboratory experiments [264]. Such holistic approaches will probably became more frequent in a near future, with the data from multi-omics technologies from GM strains being used for *in silico* modelling of genome-scale metabolic pathways that could identify limiting factors and bottlenecks, providing a second generation of targets for genetic engineering, in a virtuous circle.

4.5.3. Adaptative evolution strategies and bioprocess engineering

Following the choice of a strain and of a metabolic engineering strategy, the next steps of process development are also of major importance in the design of a successful application. They could include some optional adaptative laboratory evolution strategies, which use iterative cultivation steps under a given stress condition to obtain variants better adapted for industrial production, and the bioprocess engineering steps themselves (optimization of medium, fermentation, recovery, purification, etc.).

The use of evolutionary approaches in strain engineering requires the implementation of a suitable selection method for the product and/or phenotype of interest. The FACS methodology constitutes a preferred choice for such purposes. In one of the first examples of using adaptative laboratory evolution in *Y. lipolytica*, for riboflavin overproduction, some research teams from UT Austin and the University of California, San Francisco, compared two FACS methods for high-throughput screening: the classical single cell FACS and the more recent microdroplet-enabled FACS (droplet FACS). It appeared that single cell FACS was favoring intracellular riboflavin accumulation when droplet FACS was favoring extracellular product accumulation [266]. Beside this example, adaptative laboratory evolution strategies have recently been applied in *Y. lipolytica* for various purposes: restoring the glucose metabolism of a strain engineered for succinic acid production [267]; selecting the ionic liquid-tolerant YlCW001 strain [137]; improving limonene tolerance during its production by GM strains [268]; enhancing thermotolerance during industrial fermentation for erythritol production [269]; increasing tolerance to aromatic aldehydes or to ferulic acid, both for a more efficient lignocellulose valorization [270, 271].

Bioprocess engineering steps will not be detailed in this review, since they have been addressed more appropriately in several previous publications who highlighted, from different examples, the importance of various external factors (temperature, pH, nutrients and oxygen availability) on *Y. lipolytica* growth and productivity, together with the stress that industrial conditions can exert on yeast cells [63, 172, 272-274]. As seen in Section 2.2, this yeast benefits from a robust tolerance to a large pH range, to high salt levels and to organic solvents, all valuable assets in easing bioprocess optimization. Bioprocess engineering progresses have particularly being reviewed for heterologous proteins [172] and biodiesel production [272]. The process optimization implemented for the production of EPA-rich SCO by Dupont (cf. Section 2.3.2) has also been extensively described [46]. In the case of a dimorphic yeast like *Y. lipolytica*, the control of the dimorphic transition is an additional constraint since the cell morphology has a high impact on the production of metabolites and proteins, as reviewed previously [13, 53]. The oleaginous nature of *Y. lipolytica* presents major advantages for the bioconversion of hydrophobic substrates since the process can be performed in two-liquid biphasic systems. Thanks to its specific characteristics (naturally secreted emulsifiers, protrusions and cell surface hydrophobicity – cf. Section 2.2.5) allowing an efficient uptake of these substrates from organic solvents, GM strains expressing P450s and their reductases have been applied to the oxidation of hardly soluble hydrophobic steroids in a two-liquid biphasic system [275]. Employing such a biphasic system allowed a more efficient bioconversion compared to an aqueous system while considerably simplifying the whole process [275]. Beside their interest for the design of metabolic engineering strategies, GEMs can also be applied to predicting the metabolic responses of yeast cells to environmental conditions such as industrial-scale production, as demonstrated at the University of Graz (Austria) during the optimization of a *Y. lipolytica* bioprocess for lipid production [276]. Some statistical modeling tools, such as response surface methodology, are also now currently applied to bioprocess optimization, from wild-type as well as for GM strains [277-280]. As seen above in Section 2.3.3, a lot of research is performed on remodelling the hydrolytic secretome of *Y. lipolytica* in order to allow the use of plant biomass as a renewable and cheap carbon source, an approach aiming towards sustainable development and circular bio-economy. The final purpose is to engineer this oleaginous yeast, which natural isolates have only a limited capacity to grow on plant-derived biomass, into a valuable candidate for consolidated bioprocessing applications. Consolidated bioprocessing (CBP) requires the simultaneous production of hydrolases able to degrade plant-derived polymeric substrates and of enzymes allowing the microbial conversion of the released sugars into value-added compounds, in a single step. The progresses that have been made in this competitive research area have been reviewed in a very recent publication, in which the authors discuss the economic advantages of CBP, simulate different industrial CBP models based on GM *Y. lipolytica* and calculate the associated costs [281]. All these tools and strategies will contribute to

establish *Y. lipolytica* as a workhorse for a wide range of applications in the very competitive world of white biotechnology.

5. A conclusion in the shape of a question mark: what future for GMOs in our societies?

Since the holding of the Asilomar Conference on Recombinant DNA, in 1975 in California (USA), during which 140 biologists, physicians or law professionals discussed the potential biohazards and the need for a regulation in the nascent domain of biotechnologies, GMO have been regulated worldwide by governmental agencies [282]. This regulatory framework encompass research on GMOs, which needs to be approved by devoted committees from universities or research institutions, as well as their commercial and industrial use, including that of GMO crops and GMO-containing food. There are however some differences in the regulation between countries, especially concerning the release of GMOs, the most important ones being between the USA and European countries: the American policy is not giving as much attention to the process as other countries and takes into account a concept of substantial equivalence, when the European policy is more severe and holds to the precautionary principle [283]. Notably, the question to know if gene edited organisms (sometimes called new GMOs, including notably CRISPR-generated ones) should be concerned by the GMO regulation received different responses depending on the countries. It was a "no" for the USA, considering that no heterologous sequence was present in the resulting organism (as in the case of a mutagenesis step), but a "yes" for European countries who considered that all organisms generated through genetic engineering are indeed GMOs [284, 285]. However, this crucial question still remains presently a matter of societal debate. There are also important differences between social acceptances of GMOs in USA and in Europe, with the Europeans being notably more suspicious towards GM food and considering it more negatively than the North Americans [283]. This situation of strict regulation, especially in Europe, which contributes by the way to reinforce the social reluctance towards GMOs, constitutes a major problem for the development of GMO-based bioprocesses, above all for food applications. These problems have been highlighted in some publications that qualify the classification of GMOs of "meaningless" and this denomination of "nonsensical pseudo-category", while denouncing the "precautionary rabbit hole" of their regulation [286, 287]. If the problems of societal acceptance are particularly acute for GMO crops, this unfavourable situation also impacts microbial biotechnologies. Namely, some agro-industrial or commercial food companies may be brought to revise their strategic research choices, by favouring the use of traditionally improved strains over that of GM ones, despite the fact that these traditional technics may be less efficient and less reliable (and sometimes dangerous for the manipulator, if chemical mutagens are to be used).

To illustrate this point, the case of erythritol is interesting: its industrial production from *Y. lipolytica* by Baolingbao Biology Co. (cf. Section 2.3.1) makes use only of traditionally improved strains and the research papers that describe improvements in the domain also adopted the same strategy. Namely, as described above in Section 3.2.2, a consortium of Chinese laboratories used ARTP mutagenesis to derive the erythritol overproducer non-GM mutant M53 strain from the wild-type SWJ-1b isolate [129]. Similarly, a group of research teams from Jiangnan University used mutagenesis combined with high-throughput screening to select an erythritol overproducer non-GM mutant strain from a wild-type isolate [288]. Interestingly, some genetic engineering technics were implemented in their project, but not for the design of the production strain: these tools were used to build a genetically-encoded erythritol biosensor strain (using an erythritol-responsive transcription factor to activate expression of an eGFP-encoding gene, leading to a fluorescence signal) that was used in the high-throughput screening test applied to the mutant strains generated through a combined UV/ARTP mutagenesis step [288]. The same authors pursued the optimization of their erythritol production bioprocess by applying an adaptive

evolution strategy to the producing strain BBE-17 in order to enhance its thermotolerance during industrial fermentation, as evoked above in Section 4.5.3 [271]. Starting with a strain with an optimum temperature of 30 °C, they applied a progressive adaptive evolution scheme that allowed them to obtain, after 11 months of continuous cultivation and selection, an improved strain with an optimum temperature of 35 °C. However, the erythritol yield was considerably lower in this selected strain, which prompted the use of genetic engineering as a last resort: the authors performed a transcriptome analysis on their thermotolerant strain in order to identify the genes linked to this phenotype, reconstructed the thermotolerant phenotype using a surrogate Po1f strain and finally transferred the target gene modifications into a BBE-17Δ*Ku70*Δ*URA3* strain. This GM strain was able to produce a three-fold higher erythritol yield, without detrimental effects on cell growth, at a temperature of 33 °C compared to the parent BBE-17T strain [271]. Thus, such an example illustrates the limits of traditional mutagenesis, even when combined to adaptative evolution strategies, when used alone as an alternative to genetic engineering.

It is to hope that a future easing of the regulation policy for the new GMOs, especially for gene edited/CRISPR-generated organisms, could allow to relieve the regulatory constraints that presently limit their use in some of their numerous domains of application. Although it would be difficult to determine what influence GMO regulations and societal acceptance could have had on the strategic choices of laboratories and companies, we can note that the major food-oriented applications of *Y. lipolytica* strains (citric acid, erythritol, KGA) have majorly favoured traditionally improved strains (cf. Sections 2.3.1 and 3.2.2). If this tendency was to increase in the future, a more systematic exploration of the natural *Y. lipolytica* biodiversity for potential applications, leveraged by new mutagenesis technics (ARTP), adaptative evolution strategies and high-throughput screening technologies would constitute a valuable asset. Therefore, *Y. lipolytica* is in good position to become a biotechnological workhorse, through both traditional and genetic engineering pathways.

References

1. Groenewald, M.; Boekhout, T.; Neuvéglise, C.; Gaillardin, C.; van Dijck, P.W.; Wyss, M. *Yarrowia lipolytica*: safety assessment of an oleaginous yeast with a great industrial potential. *Crit Rev Microbiol* **2014**, *40(3)*, 187-206. doi: 10.3109/1040841X.2013.770386

2. Fröhlich-Wyder, M.T.; Arias-Roth, E.; Jakob, E. Cheese yeasts. *Yeast* **2019**, *36(3)*, 129-141. doi: 10.1002/yea.3368

3. Butinar, L.; Santos, S.; Spencer-Martins, I.; Oren, A.; Gunde-Cimerman, N. Yeast diversity in hypersaline habitats. *FEMS Microbiol Lett* **2005**, *244(2)*, 229-234. doi: 10.1016/j.femsle.2005.01.043

4. Chi, Z.; Wang, F.; Wang, L. *et al.* Selection of *Yarrowia lipolytica* strains with high protein content from yeasts isolated from different marine environments. *J Ocean Univ China* **2007** *6*, 360–364. doi: 10.1007/s11802-007-0360-7

5. Sibanda, T.; Selvarajan, R.; Tekere, M.; Nyoni, H.; Meddows-Taylor, S. Potential biotechnological capabilities of cultivable mycobiota from carwash effluents. *Microbiologyopen* 2017 *6(5)*, e00498. doi: 10.1002/mbo3.498

6. Shukla, S.P.; Plata, C.; Reichelt, M. *et al*. Microbiome-assisted carrion preservation aids larval development in a burying beetle. *Proc Natl Acad Sci U S A* **2018**, *115(44)*, 11274-11279. doi: 10.1073/pnas.1812808115

7. Desnos-Ollivier, M.; Letscher-Bru, V.; Neuvéglise, C.; Dromer, F. *Yarrowia lipolytica* causes sporadic cases and local outbreaks of infections and colonisation. *Mycoses* **2020**, *63(7)*, 737-745. doi: 10.1111/myc.13095

8. Sutherland, J.B.; Cornelison, C.; Crow, S.A. *CANDIDA | Yarrowia lipolytica (Candida lipolytica)*. In *Encyclopedia of Food Microbiology* (Second Edition); Batt, C.A.; Tortorello, M.L., Eds.; Academic Press, Cambridge, Massachusetts, USA, 2014; pp.374-378.

9. Barth, G.; Gaillardin, C. *Yarrowia lipolytica*. In *Genetics, biochemistry and molecular biology of non conventional yeasts in biotechnology*; Wolf, W.K., Ed.; Springer-Verlag, Berlin, Germany, 1996; Volume 1, pp. 313-388.

10. Barth, G.; Gaillardin, C. Physiology and genetics of the dimorphic fungus *Yarrowia lipolytica*. *FEMS Microbiol Rev* **1997**, *19(4)*, 219-237. doi: 10.1111/j.1574-6976.1997.tb00299.x

11. Thevenieau, F.; Nicaud, J.M.; Gaillardin, C. Applications of the Non-Conventional Yeast *Yarrowia lipolytica*. In *Yeast Biotechnology: Diversity and Applications*; Satyanarayana, T.; Kunze, G., Eds.; Springer, Dordrecht, 2009; pp. 589-613.

12. Pérez-Campo, F.M.; Domínguez, A. Factors affecting the morphogenetic switch in *Yarrowia lipolytica*. *Curr Microbiol* **2001**, *43(6)*, 429-433. doi: 10.1007/s002840010333

13. Timoumi, A.; Guillouet, S.E.; Molina-Jouve, C.; Fillaudeau, L.; Gorret, N. Impacts of environmental conditions on product formation and morphology of *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **2018**, *102(9)*, 3831-3848. doi: 10.1007/s00253-018-8870-3

14. Nicaud, J.M.; Fabre, E.; Gaillardin, C. Expression of invertase activity in *Yarrowia lipolytica* and its use as a selective marker. *Curr Genet* **1989**, *16(4)*, 253-260. doi: 10.1007/BF00422111

15.   Swennen, D.; Beckerich, J.M. *Yarrowia lipolytica* vesicle-mediated protein transport pathways. *BMC Evol Biol* **2007**, *7*, 219. doi: 10.1186/1471-2148-7-219

16.   Beopoulos, A.; Chardot, T.; Nicaud, J.M. *Yarrowia lipolytica*: A model and a tool to understand the mechanisms implicated in lipid accumulation. *Biochimie* **2009**, *91(6)*, 692-696. doi: 10.1016/j.biochi.2009.02.004

17.   Madzak, C.; Gaillardin, C.; Beckerich, J.M. Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review. *J Biotechnol* **2004**, *109(1-2)*, 63-81. doi: 10.1016/j.jbiotec.2003.10.027

18.   Madzak, C.; Beckerich, J.M. Heterologous Protein Expression and Secretion in *Yarrowia lipolytica*. In *Yarrowia lipolytica*; Barth, G., Ed.; Springer, Berlin, Heidelberg, Germany, 2013; Microbiology Monographs, Volume 25, pp. 1-76.

19.   Madzak, C. *Yarrowia lipolytica*: recent achievements in heterologous protein expression and pathway engineering. *Appl Microbiol Biotechnol* **2015**, *99(11)*, 4559-4577. doi: 10.1007/s00253-015-6624-z

20.   De Pourcq, K.; Vervecken, W.; Dewerte, I.; Valevska, A.; Van Hecke, A.; Callewaert, N. Engineering the yeast *Yarrowia lipolytica* for the production of therapeutic proteins homogeneously glycosylated with $Man_8GlcNAc_2$ and $Man_5GlcNAc_2$. *Microb Cell Fact* **2012**, *11*, 53. doi: 10.1186/1475-2859-11-53

21.   Kim, H.; Yoo, S.J.; Kang, H.A. Yeast synthetic biology for the production of recombinant therapeutic proteins. *FEMS Yeast Res* **2015**, *15(1)*, 1-16. doi: 10.1111/1567-1364.12195

22.   Fickers, P.; Benetti, P.H.; Waché, Y. *et al*. Hydrophobic substrate utilisation by the yeast *Yarrowia lipolytica*, and its potential applications. *FEMS Yeast Res* **2005**, *5(6-7)*, 527-543. doi: 10.1016/j.femsyr.2004.09.004

23.   Lasserre, J.P.; Nicaud, J.M.; Pagot, Y.; Joubert-Caron, R.; Caron, M.; Hardouin, J. First complexomic study of alkane-binding protein complexes in the yeast *Yarrowia lipolytica*. *Talanta* **2010**, *80(4)*, 1576-1585. doi: 10.1016/j.talanta.2009.07.016

24.   Liu, H.; Song, Y.; Fan, X.; Wang, C.; Lu, X.; Tian, Y. *Yarrowia lipolytica* as an Oleaginous Platform for the Production of Value-Added Fatty Acid-Based Bioproducts. *Front Microbiol* **2021**, *11*, 608662. doi: 10.3389/fmicb.2020.608662

25.   Casarégola, S.; Feynerol, C.; Diez, M.; Fournier, P.; Gaillardin, C. Genomic organization of the yeast *Yarrowia lipolytica*. *Chromosoma* **1997**, *106(6)*, 380-390. doi: 10.1007/s004120050259

26.   Dujon, B.; Sherman, D.; Fischer, G. *et al*. Genome evolution in yeasts. *Nature* **2004**, *430(6995)*, 35-44. doi: 10.1038/nature02579

27.   Neuvéglise, C.; Marck, C.; Gaillardin, C. The intronome of budding yeasts. *C R Biol* **2011**, *334(8-9)*, 662-670. doi: 10.1016/j.crvi.2011.05.015

28.   Magnan, C.; Yu, J.; Chang, I. *et al*. Sequence Assembly of *Yarrowia lipolytica* Strain W29/CLIB89 Shows Transposable Element Diversity. *PLoS One* **2016**, *11(9)*, e0162363. doi: 10.1371/journal.pone.0162363

29.   Devillers, H.; Neuvéglise, C. Genome Sequence of the Oleaginous Yeast *Yarrowia lipolytica* H222. *Microbiol Resour Announc* **2019**, *8(4)*, e01547-18. doi: 10.1128/MRA.01547-18

30.   Casaregola, S.; Neuvéglise, C.; Lépingle, A. *et al*. Genomic exploration of the hemiascomycetous yeasts: 17. *Yarrowia lipolytica*. *FEBS Lett* **2000**, *487(1)*, 95-100. doi: 10.1016/s0014-5793(00)02288-2

31.   Schmid-Berger, N.; Schmid, B.; Barth G. Ylt1, a highly repetitive retrotransposon in the genome of the dimorphic fungus *Yarrowia lipolytica*. *J Bacteriol* **1994**, *176(9)*, 2477-2482. doi: 10.1128/jb.176.9.2477-2482.1994

32.   Casaregola, S.; Neuvéglise, C.; Bon, E.; Gaillardin, C. Ylli, a non-LTR retrotransposon L1 family in the dimorphic yeast *Yarrowia lipolytica*. *Mol Biol Evol* **2002**, *19(5)*, 664-677. doi: 10.1093/oxfordjournals.molbev.a004125

33.   Kovalchuk, A.; Senam, S.; Mauersberger, S.; Barth, G. Tyl6, a novel Ty3/gypsy-like retrotransposon in the genome of the dimorphic fungus *Yarrowia lipolytica*. *Yeast* **2005**, *22(12)*, 979-991. doi: 10.1002/yea.1287

34.   Nicaud, J.M.; Madzak, C.; van den Broek, P. *et al*. Protein expression and secretion in the yeast *Yarrowia lipolytica*. *FEMS Yeast Res* **2002**, *2(3)*, 371-379. doi: 10.1016/S1567-1356(02)00082-X

35.   Madzak, C. New tools for heterologous protein production in the yeast *Yarrowia lipolytica*. In *Recent research developments in microbiology*; Pandalai, S.G., Ed.; Research Signpost, Trivandrum, India, 2003; Volume 7, pp. 453-479.

36.   Bankar, A.V.; Kumar, A.R.; Zinjarde, S.S. Environmental and industrial applications of *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **2009**, *84(5)*, 847-865. doi: 10.1007/s00253-009-2156-8

37.   Sibirny, A.; Madzak, C.; Fickers, P. Genetic engineering of non-conventional yeast for the production of valuable compounds. In *Microbial biotechnology: Progress and trends*; Harzevili, F.D.; Chen, H., Eds.; CRC Press, Boca Raton, Florida, USA, 2014; pp. 63-111.

38.   Zinjarde, S.S. Food-related applications of *Yarrowia lipolytica*. *Food Chem* **2014**, *152*, 1-10. doi: 10.1016/j.foodchem.2013.11.117

39.   Żogała, B.; Robak, M.; Rymowicz, W.; Wzientek, K.; Rusin, M.; Maruszczak, J. Geoelectrical observation of *Yarrowia lipolytica* bioremediation of petrol contaminated soil. *Pol J Environ Stud* **2005**, *14*, 665-669.

40.   Robak, M.; Boruczkowski, T.; Drozdz, W.; Lazar, Z.; Baranowska, M.; Przado, D. Application of the yeasts Yarrowia lipolytica for in-situ bioremediation of soil contaminated with creosote oil—a case study. *Ochrona Srodowiska* **2011**, *33*, 27-33.

41.   Fickers, P.; Nicaud, J.M.; Destain, J.; Thonart, P. Overproduction of lipase by *Yarrowia lipolytica* mutants. *Appl Microbiol Biotechnol* **2003**, *63(2)*, 136-142. doi: 10.1007/s00253-003-1342-3

42.   Madzak, C. *Yarrowia lipolytica* engineering as a source of microbial cell factories. In *Microbial Cell Factories Engineering for Production of Biomolecules*; Singh, V., Ed.; Academic Press, Cambridge, Massachusetts, USA, 2021; pp 345-380.

43.   Madzak, C. Engineering *Yarrowia lipolytica* for Use in Biotechnological Applications: A Review of Major Achievements and Recent Innovations. *Mol Biotechnol* **2018**, *60(8)*, 621-635. doi: 10.1007/s12033-018-0093-4

44.   Abdel-Mawgoud, A.M.; Markham, K.A.; Palmer, C.M.; Liu, N.; Stephanopoulos, G.; Alper, H.S. Metabolic engineering in the host *Yarrowia lipolytica*. *Metab Eng* **2018**, *50*, 192-208. doi: 10.1016/j.ymben.2018.07.016

45. Xue, Z.; Sharpe, P.L.; Hong, S.P. et al. Production of omega-3 eicosapentaenoic acid by metabolic engineering of *Yarrowia lipolytica*. *Nat Biotechnol* **2013**, *31(8)*, 734-740. doi: 10.1038/nbt.2622

46. Xie, D.; Jackson, E.N.; Zhu, Q. Sustainable source of omega-3 eicosapentaenoic acid from metabolically engineered *Yarrowia lipolytica*: from fundamental research to commercial production. *Appl Microbiol Biotechnol* **2015**, *99(4)*, 1599-1610. doi: 10.1007/s00253-014-6318-y

47. Hollands, K.; Baron, C.M.; Gibson, K.J. *et al*. Engineering two species of yeast as cell factories for 2'-fucosyllactose. *Metab Eng* **2019**, *52*, 232-242. doi: 10.1016/j.ymben.2018.12.005

48. Pignède, G.; Wang, H.J.; Fudalej, F.; Seman, M.; Gaillardin, C.; Nicaud, J.M. Autocloning and amplification of *LIP2* in *Yarrowia lipolytica*. *Appl Environ Microbiol* **2000**, *66(8)*, 3283-3289. doi: 10.1128/aem.66.8.3283-3289.2000

49. Tiels, P.; Baranova, E.; Piens, K. *et al*. A bacterial glycosidase enables mannose-6-phosphate modification and improved cellular uptake of yeast-produced recombinant human lysosomal enzymes. *Nat Biotechnol* **2012**, *30(12)*, 1225-1231. doi: 10.1038/nbt.2427

50. Ledesma-Amaro, R.; Nicaud, J.M. Metabolic Engineering for Expanding the Substrate Range of *Yarrowia lipolytica*. *Trends Biotechnol* **2016**, *34(10)*, 798-809. doi: 10.1016/j.tibtech.2016.04.010

51. Spagnuolo, M.; Shabbir Hussain, M.; Gambill, L.; Blenner, M. Alternative Substrate Metabolism in *Yarrowia lipolytica*. *Front Microbiol* **2018**, *9*, 1077. Published 2018 May 25. doi: 10.3389/fmicb.2018.01077

52. Miller, K.K.; Alper, H.S. *Yarrowia lipolytica*: more than an oleaginous workhorse. *Appl Microbiol Biotechnol* **2019**, *103(23-24)*, 9251-9262. doi: 10.1007/s00253-019-10200-x

53. Soong, Y.V.; Liu, N.; Yoon, S.; Lawton, C.; Xie, D. Cellular and metabolic engineering of oleaginous yeast *Yarrowia lipolytica* for bioconversion of hydrophobic substrates into high-value products. *Eng Life Sci* **2019**, *19(6)*, 423-443. doi: 10.1002/elsc.201800147

54. Hapeta, P.; Rakicka, M.; Dulermo, R. *et al*. Transforming sugars into fat - lipid biosynthesis using different sugars in *Yarrowia lipolytica*. *Yeast* **2017**, *34(7)*, 293-304. doi: 10.1002/yea.3232

55. Niehus, X.; Crutz-Le Coq, A.M.; Sandoval, G.; Nicaud, J.M.; Ledesma-Amaro, R. Engineering *Yarrowia lipolytica* to enhance lipid production from lignocellulosic materials. *Biotechnol Biofuels* **2018**, *11*, 11. doi: 10.1186/s13068-018-1010-6

56. Ryu, S.; Hipp, J.; Trinh, C.T. Activating and Elucidating Metabolism of Complex Sugars in *Yarrowia lipolytica*. *Appl Environ Microbiol* **2015**, *82(4)*, 1334-1345. doi: 10.1128/AEM.03582-15

57. Li, H.; Alper, H.S. Enabling xylose utilization in *Yarrowia lipolytica* for lipid production. *Biotechnol J* **2016**, *11(9)*, 1230-1240. doi: 10.1002/biot.201600210

58. Li, H.; Alper, H.S. Producing Biochemicals in *Yarrowia lipolytica* from Xylose through a Strain Mating Approach. *Biotechnol J* **2020** *15(2)*, e1900304. doi: 10.1002/biot.201900304

59. Guo, H.; Su, S.; Madzak, C.; Zhou, J.; Chen, H.; Chen, G. Applying pathway engineering to enhance production of alpha-ketoglutarate in *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **2016**, *100(23)*, 9875-9884. doi: 10.1007/s00253-016-7913-x

60. Braga, A.; Belo, I. Biotechnological production of γ-decalactone, a peach like aroma, by *Yarrowia lipolytica*. *World J Microbiol Biotechnol* **2016**, *32(10)*, 169. doi: 10.1007/s11274-016-2116-2

61. Yan, J.; Han, B.; Gui, X. *et al*. Engineering *Yarrowia lipolytica* to Simultaneously Produce Lipase and Single Cell Protein from Agro-industrial Wastes for Feed. *Sci Rep* **2018**, *8(1)*, 758. doi: 10.1038/s41598-018-19238-9

62. Shi, N.; Mao, W.; He, X.; Chi, Z.; Chi, Z.; Liu, G. Co-expression of Exo-inulinase and Endo-inulinase Genes in the Oleaginous Yeast *Yarrowia lipolytica* for Efficient Single Cell Oil Production from Inulin. *Appl Biochem Biotechnol* **2018**, *185(1)*, 334-346. doi: 10.1007/s12010-017-2659-1

63. Worland, A.M.; Czajka, J.J.; Li, Y.; Wang, Y.; Tang, Y.J.; Su, W.W. Biosynthesis of terpene compounds using the non-model yeast *Yarrowia lipolytica*: grand challenges and a few perspectives. *Curr Opin Biotechnol* **2020**, *64*, 134-140. doi: 10.1016/j.copbio.2020.02.020

64. Muhammad, A.; Feng, X.; Rasool, A.; Sun, W.; Li, C. Production of plant natural products through engineered *Yarrowia lipolytica*. *Biotechnol Adv* **2020**, *43*, 107555. doi: 10.1016/j.biotechadv.2020.107555

65. Li, Z.J.; Wang, Y.Z.; Wang, L.R.; Shi, T.Q.; Sun, X.M.; Huang, H. Advanced Strategies for the Synthesis of Terpenoids in *Yarrowia lipolytica*. *J Agric Food Chem* **2021**, *69(8)*, 2367-2381. doi: 10.1021/acs.jafc.1c00350

66. Tong, Y.; Zhou, J.; Zhang, L.; Xu, P. A Golden-Gate Based Cloning Toolkit to Build Violacein Pathway Libraries in *Yarrowia lipolytica*. *ACS Synth Biol* **2021**, *10(1)*, 115-124. doi: 10.1021/acssynbio.0c00469

67. Fickers, P.; Cheng, H.; Sze Ki Lin, C. Sugar Alcohols and Organic Acids Synthesis in *Yarrowia lipolytica*: Where Are We? *Microorganisms* **2020**, *8(4)*, 574. doi: 10.3390/microorganisms8040574

68. Abbasi, A.R.; Liu, J.; Wang, Z.; Zhao, A.; Ying, H.; Qu, L.; Alam, M.A.; Xiong, W.; Xu, J.; Lv, Y. Recent Advances in Producing Sugar Alcohols and Functional Sugars by Engineering *Yarrowia lipolytica*. *Front. Bioeng. Biotechnol* **2021**, *9*, 648382. doi: 10.3389/fbioe.2021.648382

69. Gao, Q.; Yang, J.L.; Zhao, X.R. *et al*. *Yarrowia lipolytica* as a Metabolic Engineering Platform for the Production of Very-Long-Chain Wax Esters. *J Agric Food Chem* **2020**, *68(39)*, 10730-10740. doi: 10.1021/acs.jafc.0c04393

70. Gao, C.; Qi, Q.; Madzak, C.; Lin, C.S. Exploring medium-chain-length polyhydroxyalkanoates production in the engineered yeast *Yarrowia lipolytica*. *J Ind Microbiol Biotechnol* **2015** *42(9)*, 1255-1262. doi: 10.1007/s10295-015-1649-y

71. Lajus, S.; Dusséaux, S.; Verbeke, J. *et al*. Engineering the Yeast *Yarrowia lipolytica* for Production of Polylactic Acid Homopolymer. *Front Bioeng Biotechnol* **2020**, *8*, 954. doi: 10.3389/fbioe.2020.00954

72. Wang, L.; Zong, Z.; Liu, Y. *et al*. Metabolic engineering of *Yarrowia lipolytica* for the biosynthesis of crotonic acid. *Bioresour Technol* **2019**, *287*, 121484. doi: 10.1016/j.biortech.2019.121484

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

74. Abghari, A.; Madzak, C.; Chen, S. Combinatorial Engineering of *Yarrowia lipolytica* as a Promising Cell Biorefinery Platform for the de novo Production of Multi-Purpose Long Chain Dicarboxylic Acids. *Fermentation* **2017**, *3*, 40.

75. Yan, J.; Yan, Y.; Madzak, C.; Han, B. Harnessing biodiesel-producing microbes: from genetic engineering of lipase to metabolic engineering of fatty acid biosynthetic pathway. *Crit Rev Biotechnol* **2017**, *37(1)*, 26-36. doi: 10.3109/07388551.2015.1104531

76. Spagnuolo, M.; Yaguchi, A.; Blenner, M. Oleaginous yeast for biofuel and oleochemical production. *Curr Opin Biotechnol* **2019**, *57*, 73-81. doi: 10.1016/j.copbio.2019.02.011

77. Ma, J.; Gu, Y.; Marsafari, M.; Xu, P. Synthetic biology, systems biology, and metabolic engineering of *Yarrowia lipolytica* toward a sustainable biorefinery platform. *J Ind Microbiol Biotechnol* **2020**, *47(9-10)*, 845-862. doi: 10.1007/s10295-020-02290-8

78. Liu, Z.; Moradi, H.; Shi, S.; Darvishi, F. Yeasts as microbial cell factories for sustainable production of biofuels. *Renew Sustain Energy Rev* **2021**, *143*, 110907. doi: 10.1016/j.rser.2021.110907

79. Blazeck, J.; Liu, L.; Knight, R.; Alper, H.S. Heterologous production of pentane in the oleaginous yeast *Yarrowia lipolytica*. *J Biotechnol* **2013**, *165(3-4)*, 184-194. doi: 10.1016/j.jbiotec.2013.04.003

80. Bruder, S.; Moldenhauer, E.J.; Lemke, R.D.; Ledesma-Amaro, R.; Kabisch, J. Drop-in biofuel production using fatty acid photodecarboxylase from *Chlorella variabilis* in the oleaginous yeast *Yarrowia lipolytica*. *Biotechnol Biofuels* **2019**, *12*, 202. doi: 10.1186/s13068-019-1542-4

81. Li, J.; Ma, Y.; Liu, N. *et al*. Synthesis of high-titer alka(e)nes in *Yarrowia lipolytica* is enabled by a discovered mechanism. *Nat Commun* **2020**, *11(1)*, 6198. doi: 10.1038/s41467-020-19995-0

82. Xu, P.; Qiao, K.; Ahn, W.S.; Stephanopoulos, G. Engineering *Yarrowia lipolytica* as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proc Natl Acad Sci USA* **2016**, *113(39)*, 10848-10853. doi: 10.1073/pnas.1607295113

83. Yang, K.; Qiao, Y.; Li, F. *et al*. Subcellular engineering of lipase dependent pathways directed towards lipid related organelles for highly effectively compartmentalized biosynthesis of triacylglycerol derived products in *Yarrowia lipolytica*. *Metab Eng* **2019**, *55*, 231-238. doi: 10.1016/j.ymben.2019.08.001

84. Yang, K.; Li, F.; Qiao, Y.; Zhou, Q.; Hu, Z.; He, Y.; Yan, Y.; Xu, L.; Madzak, C.; Yan, J. Design of a New Multienzyme Complex Synthesis System Based on *Yarrowia lipolytica* Simultaneously Secreted and Surface Displayed Fusion Proteins for Sustainable Production of Fatty Acid-Derived Hydrocarbons. *ACS Sustainable Chem Eng* **2018**, *6(12)*, 17035–17043. doi: 10.1021/acssuschemeng.8b04401

85. Pimprikar, P.S.; Joshi, S.S.; Kumar, A.R.; Zinjarde, S.S.; Kulkarni, S.K. Influence of biomass and gold salt concentration on nanoparticle synthesis by the tropical marine yeast *Yarrowia lipolytica* NCIM 3589. *Colloids Surf B Biointerfaces* **2009**, *74(1)*, 309-316. doi:10.1016/j.colsurfb.2009.07.040

86. Ben Tahar, I.; Fickers, P.; Dziedzic, A.; Płoch, D.; Skóra, B.; Kus-Liśkiewicz, M. Green pyomelanin-mediated synthesis of gold nanoparticles: modelling and design, physico-chemical and biological characteristics. *Microb Cell Fact* **2019**, *18(1)*, 210. doi: 10.1186/s12934-019-1254-2

87. Han, Z.; Madzak, C.; Su, W.W. Tunable nano-oleosomes derived from engineered *Yarrowia lipolytica*. *Biotechnol Bioeng* **2013**, *110(3)*, 702-710. doi: 10.1002/bit.24761

88. Wang, H.; Wang, Z.; Liu, G.; Cheng, X.; Chi, Z.; Madzak, C.; Liu, C.; Chi, Z. Genetical Surface Display of Silicatein on *Yarrowia lipolytica* Confers Living and Renewable Biosilica-Yeast Hybrid Materials. *ACS Omega* **2020**, *5*, 7555-7566. doi: 10.1021/acsomega.0c00393

89. Tang, S.; Xi, W.; Cheng, Z. *et al*. A Living Eukaryotic Autocementation Kit from Surface Display of Silica Binding Peptides on *Yarrowia lipolytica*. *ACS Synth Biol* **2016**, *5(12)*, 1466-1474. doi: 10.1021/acssynbio.6b00085

90. Péter, G.; Dlauchy, D.; Vasdinyei, R.; Tornai-Lehoczki, J.; Deák, T. *Candida galli* sp. nov., a new yeast from poultry. *Antonie Van Leeuwenhoek* **2004**, *86(2)*, 105-110. doi: 10.1023/B:ANTO.0000036117.03847.bd

91. Knutsen, A.K.; Robert, V.; Poot, G.A. *et al*. Polyphasic re-examination of *Yarrowia lipolytica* strains and the description of three novel *Candida* species: *Candida oslonensis* sp. nov., *Candida alimentaria* sp. nov. and *Candida hollandica* sp. nov. *Int J Syst Evol Microbiol* **2007**, *57(Pt 10)*, 2426-2435. doi: 10.1099/ijs.0.65200-0

92. Limtong, S.; Youngmanitchai, W.; Kawasaki, H.; Seki, T. *Candida phangngensis* sp. nov., an anamorphic yeast species in the *Yarrowia* clade, isolated from water in mangrove forests in Phang-Nga Province, Thailand *Int J Syst Evol Microbiol* **2008**, *58(Pt 2)*, 515-519. doi: 10.1099/ijs.0.65506-0

93. Chang, C.F.; Chen, C.C.; Lee, C.F.; Liu, S.M. Identifying and characterizing *Yarrowia keelungensis* sp. nov., an oil-degrading yeast isolated from the sea surface microlayer. *Antonie Van Leeuwenhoek* **2013**, *104(6)*, 1117-1123. doi: 10.1007/s10482-013-0033-z

94. Groenewald, M.; Smith, M.T. The teleomorph state of *Candida deformans* Langeron & Guerra and description of *Yarrowia yakushimensis* comb. nov. *Antonie Van Leeuwenhoek* **2013**, 103(5), 1023-1028. doi: 10.1007/s10482-013-9882-8

95. Nagy E, Niss M, Dlauchy D, Arneborg N, Nielsen DS, Péter G. *Yarrowia divulgata* f.a., sp. nov., a yeast species from animal-related and marine sources. *Int J Syst Evol Microbiol* **2013**, *63(Pt 12)*, 4818-4823. doi: 10.1099/ijs.0.057208-0

96. Nagy, E.; Dlauchy, D.; Medeiros, A.O.; Péter, G.; Rosa, C.A. *Yarrowia porcina* sp. nov. and *Yarrowia bubula* f.a. sp. nov., two yeast species from meat and river sediment. *Antonie Van Leeuwenhoek* **2014**, *105(4)*, 697-707. doi: 10.1007/s10482-014-0125-4

97. Bing, J.; You, Z.; Zheng, Q.; Tang, J.; Ran, Y.; Huang, G. Biological and genomic analyses of a clinical isolate of *Yarrowia galli* from China. *Curr Genet* **2020**, *66(3)*, 549-559. doi: 10.1007/s00294-019-01046-x

98. Liu, K.F.; Li, X.H.; Hui, F.L. *Yarrowia brassicae* f.a., sp. nov., a new yeast species from traditional Chinese sauerkraut. *Int J Syst Evol Microbiol* **2018**, *68(6)*, 2024-2027. doi: 10.1099/ijsem.0.002783

99.   Rakicka, M.; Kieroń, A.; Hapeta, P.; Neuvéglise, C.; Lazar, Z. Sweet and sour potential of yeast from the *Yarrowia* clade. *Biomass Bioenergy* **2016**, *92*, 48-54. doi: 10.1016/j.biombioe.2016.06.0040961-9534

100. Quarterman, J.; Slininger, P.J.; Kurtzman, C.P.; Thompson, S.R.; Dien, B.S. A survey of yeast from the *Yarrowia* clade for lipid production in dilute acid pretreated lignocellulosic biomass hydrolysate. *Appl Microbiol Biotechnol* **2017**, *101(8)*, 3319-3334. doi: 10.1007/s00253-016-8062-y

101. Morin, N.; Czerwiec, Q.; Nicaud, J.M.; Neuvéglise, C.; Rossignol, T. Transforming *Candida hispaniensis*, a promising oleaginous and flavogenic yeast. *Yeast* **2020**, *37(5-6)*, 348-355. doi: 10.1002/yea.3466

102. Červenák, F.; Juríková, K.; Devillers, H. *et al*. Identification of telomerase RNAs in species of the *Yarrowia* clade provides insights into the co-evolution of telomerase, telomeric repeats and telomere-binding proteins. *Sci Rep* **2019**, *9(1)*, 13365. doi: 10.1038/s41598-019-49628-6

103. Michely, S.; Gaillardin, C.; Nicaud, J.M.; Neuvéglise, C. Comparative physiology of oleaginous species from the *Yarrowia* clade. *PLoS One* **2013**, *8(5)*, e63356. doi: 10.1371/journal.pone.0063356

104. Quarterman, J.C.; Slininger, P.J.; Hector, R.E.; Dien, B.S. Engineering *Candida phangngensis*-an oleaginous yeast from the *Yarrowia* clade-for enhanced detoxification of lignocellulose-derived inhibitors and lipid overproduction. *FEMS Yeast Res* **2018**, *18(8)*, 10.1093/femsyr/foy102. doi: 10.1093/femsyr/foy102

105. Wojtatowicz, M.; Rymowicz, W.; Kautola, H. Comparison of different strains of the yeast *Yarrowia lipolytica* for citric acid production from glucose hydrol. *Appl Biochem Biotechnol* **1991**, *31*, 165–174. doi: 10.1007/BF02921787

106. Rywińska, A.; Rymowicz, W.; Zarowska, B.; Skrzypiński, A. Comparison of citric acid production from glycerol and glucose by different strains of *Yarrowia lipolytica*. *World J Microbiol Biotechnol* **2010**, *26*, 1217-1224. doi: 10.1007/s11274-009-0291-0

107. Devillers, H.; Brunel, F.; Połomska, X.; Sarilar, V.; Lazar, Z.; Robak, M.; Neuvéglise, C. Draft genome sequence of *Yarrowia lipolytica* strain A-101 isolated from polluted soil in Poland. *Genome Announc* **2016**, *4(5)*, e01094-16. doi: 10.1128/genomeA.01094-16

108. Lazar, Z.; Rossignol, T.; Verbeke, J.; Crutz-Le Coq, A.M.; Nicaud, J.M.; Robak, M. Optimized invertase expression and secretion cassette for improving *Yarrowia lipolytica* growth on sucrose for industrial applications. *J Ind Microbiol Biotechnol* **2013**, *40*, 1273-1283. doi: 10.1007/s10295-013-1323-1

109. Rywińska, A.; Juszczyk, P.; Wojtatowicz, M.; Robak, M.; Lazar, Z.; Tomaszewska, L.; Rymowicz, W. Glycerol as a promising substrate for *Yarrowia lipolytica* biotechnological applications. *Biomass Bioenerg* **2013**, *48*, 148-166. doi: 10.1016/j.biombioe.2012.11.021

110. Papanikolaou, S.; Galiotou-Panayotou, M.; Chevalot, I.; Komaitis, M.; Marc, I.; Aggelis, G. Influence of glucose and saturated free-fatty acid mixtures on citric acid and lipid production by *Yarrowia lipolytica*. *Curr Microbiol* **2006**, *52(2)*, 134-142. doi: 10.1007/s00284-005-0223-7

111. Papanikolaou, S.; Galiotou-Panayotou, M.; Fakas, S.; Komaitis, M.; Aggelis, G. Citric acid production by *Yarrowia lipolytica* cultivated on olive-mill wastewater-based media. *Bioresour Technol* **2008**, *99(7)*, 2419-2428. doi: 10.1016/j.biortech.2007.05.005

112. Zhao, C.H.; Cui, W.; Liu, X.Y.; Chi, Z.M.; Madzak, C. Expression of inulinase gene in the oleaginous yeast *Yarrowia lipolytica* and single cell oil production from inulin-containing materials. *Metab Eng* **2010**, *12(6)*, 510-517. doi: 10.1016/j.ymben.2010.09.001

113. Wang, G.Y.; Zhang, Y.; Chi, Z.; Liu, G.L.; Wang, Z.P.; Chi, Z.M. Role of pyruvate carboxylase in accumulation of intracellular lipid of the oleaginous yeast *Yarrowia lipolytica* ACA-DC 50109. *Appl Microbiol Biotechnol* **2015**, *99(4)*, 1637-1645. doi: 10.1007/s00253-014-6236-z

114. von Arx, J.A. On endomyces, Endomycopsis and related yeast-like fungi. *Antonie Van Leeuwenhoek* **1972**, *38(3)*, 289-309. doi: 10.1007/BF02328100

115. Dos Santos, E.O.; Michelon, M.; Furlong, E.B.; Burkert, J.F.; Kalil, S.J.; Burkert, C.A. Evaluation of the composition of culture medium for yeast biomass production using raw glycerol from biodiesel synthesis. *Braz J Microbiol* **2012**, *43(2)*, 432-440. doi: 10.1590/S1517-83822012000200002

116. Yamada, K.; Machida, H. Studies on the production of lipase by microorganisms. Part IV. Identification of a new strain and determination of its lipase activity. *Nippon Nogeikagaku Kaishi* **1963**, *37*, 649-652. doi: 10.1271/nogeikagaku1924.37.649

117. Briffaud, J.; Engasser, J.M. Citric acid production from glucose. I. Growth and excretion kinetics in a stirred fermentor. *Biotechnol Bioeng* **1979**, *21*, 2083-2092. doi: 10.1002/bit.260211113

118. Briffaud J, Engasser JM. Citric acid production from glucose. II. Growth and excretion kinetics in a trickle-flow fermentor. *Biotechnol Bioeng* **1979**, *21*, 2093-2111. doi: 10.1002/bit.260211114

119. Wentworth, S.D.; Cooper, D.G. Self-cycling fermentation of a citric acid producing strain of *Candida lipolytica*. *J Ferment Bioeng* **1996**, *81*, 400-405. doi: 10.1016/0922-338X(96)85140-3

120. Barth, G.; Weber, H. Genetic studies in the yeast *Saccharomyces lipolytica*: Inactivation and mutagenesis. *Z Allg Mikrobiol* **1983**, *23*, 147-157. doi: 10.1002/jobm.3630230302

121. Förster, A.; Aurich, A.; Mauersberger, S.; Barth, G. Citric acid production from sucrose using a recombinant strain of the yeast *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **2007**, *75(6)*, 1409-1417. doi: 10.1007/s00253-007-0958-0

122. Moeller, L.; Zehnsdorf, A.; Aurich, A.; Bley, T.; Strehlitz, B. Substrate utilization by recombinant *Yarrowia lipolytica* growing on sucrose. *Appl Microbiol Biotechnol* **2012**, *93(4)*, 1695-1702. doi: 10.1007/s00253-011-3681-9

123. Zinjarde, S.S.; Pant, A.A. Hydrocarbon degraders from tropical marine environments. *Mar Pollut Bull* **2002**, *44(2)*, 118-121. doi:10.1016/s0025-326x(01)00185-0

124. Dusane, D.H.; Nancharaiah, Y.V.; Venugopalan, V.P.; Kumar, A.R.; Zinjarde, S.S. Biofilm formation by a biotechnologically important tropical marine yeast isolate, *Yarrowia lipolytica* NCIM 3589. *Water Sci Technol* **2008**, *58(12)*, 2467-2475. doi:10.2166/wst.2008.585

125. Zinjarde, S.S.; Pant, A. Emulsifier from a tropical marine yeast, *yarrowia lipolytica* NCIM 3589. *J Basic Microbiol* **2002**, *42(1)*, 67-73. doi:10.1002/1521-4028(200203)42:1<67::AID-JOBM67>3.0.CO;2-M

126. Liu, X.; Wang, X.; Xu, J.; Xia, J. *et al*. Citric acid production by *Yarrowia lipolytica* SWJ-1b using corn steep liquor as a source of organic nitrogen and vitamins. *Ind Crop Prod* **2015**, *78*, 154-160. doi: 10.1016/j.indcrop.2015.10.029

127. Tan, M.J.; Chen, X.; Wang, Y.K.; Liu, G.L.; Chi, Z.M. Enhanced citric acid production by a yeast *Yarrowia lipolytica* over-expressing a pyruvate carboxylase gene. *Bioprocess Biosyst Eng* **2016**, *39(8)*, 1289-96. doi: 10.1007/s00449-016-1607-8

128. Cui, W.; Wang, Q.; Zhang, F.; Zhang, S.C.; Chi, Z.M.; Madzak, C. Direct conversion of inulin into single cell protein by the engineered *Yarrowia lipolytica* carrying inulinase gene. *Process Biochem* **2011**, *46(7)*, 1442-1448. doi: 10.1016/j.procbio.2011.03.017

129. Liu, X.; Lv ; J.; Xu, J. *et al*. Erythritol production by *Yarrowia lipolytica* mutant strain M53 generated through atmospheric and room temperature plasma mutagenesis. *Food Sci Biotechnol* **2017**, *26(4)*, 979-986. doi:10.1007/s10068-017-0116-0

130. Gaillardin, C.M.; Charoy, V.; Heslot, H. A study of copulation, sporulation and meiotic segregation in *Candida lipolytica*. *Arch Mikrobiol* **1973**, *92*, 69-83. doi: 10.1007/BF00409513

131. Ogrydziak, D.M.; Nicaud, J.M. Characterization of *Yarrowia lipolytica XPR2* multi-copy strains over-producing alkaline extracellular protease - a system for rapidly increasing secretory pathway cargo loads. *FEMS Yeast Res* **2012**, *12(8)*, 938-948. doi: 10.1111/j.1567-1364.2012.00846.x

132. Pomraning, K.R.; Baker, S.E. Draft genome sequence of the dimorphic yeast *Yarrowia lipolytica* strain W29. *Genome Announc* **2015**, *3*, e01211-15. doi: 10.1128/genomeA.01211-15

133. Wickerham, L.J.; Kurtzman, C.P.; Herman, A.I. Sexual reproduction in *Candida lipolytica Science* **1970**, *167(3921)*, 1141. doi:10.1126/science.167.3921.1141

134. Le Dall, M.T.; Nicaud, J.M.; Gaillardin, C. Multiple-copy integration in the yeast *Yarrowia lipolytica*. *Curr Genet* **1994**, *26(1)*, 38-44. doi: 10.1007/BF00326302

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

136. Liu, L.; Alper, H.S. Draft Genome Sequence of the Oleaginous Yeast *Yarrowia lipolytica* PO1f, a Commonly Used Metabolic Engineering Host. *Genome Announc* **2014**, *2(4)*, e00652-14. doi: 10.1128/genomeA.00652-14

137. Walker, C.; Ryu, S.; Haridas, S. *et al*. Draft Genome Assemblies of Ionic Liquid-Resistant *Yarrowia lipolytica* PO1f and Its Superior Evolved Strain, YlCW001. *Microbiol Resour Announc* **2020**, *9(9)*, e01356-19. doi: 10.1128/MRA.01356-19

138. Leplat, C.; Nicaud, J.M.; Rossignol, T. High-throughput transformation method for *Yarrowia lipolytica* mutant library screening. *FEMS Yeast Res* **2015**, *15(6)*, fov052. doi:10.1093/femsyr/fov052

139. Zhou, J.W.; Zhou, H.Y.; Du, G.C.; Liu, L.M.; Chen, J. Screening of a thiamine-auxotrophic yeast for alpha-ketoglutaric acid overproduction. *Lett Appl Microbiol* **2010**, *51(3)*, 264–271. doi: 10.1111/j.1472-765X.2010.02889.x

140. Zeng, W.; Zhang, H.; Xu, S.; Fang, F.; Zhou, J. Biosynthesis of keto acids by fed-batch culture of *Yarrowia lipolytica* WSH-Z06. *Bioresour Technol* **2017**, *243*, 1037-1043. doi: 10.1016/j.biortech.2017.07.063

141. Zeng, W.; Fang, F.; Liu, S.; Du, G.; Chen, J.; Zhou, J. Comparative genomics analysis of a series of *Yarrowia lipolytica* WSH-Z06 mutants with varied capacity for α-ketoglutarate production. *J Biotechnol* **2016**, *239*, 76-82. doi: 10.1016/j.jbiotec.2016.10.008

142. Yin, X.; Madzak, C.; Du, G.; Zhou, J.; Chen, J. Enhanced alpha-ketoglutaric acid production in *Yarrowia lipolytica* WSH-Z06 by regulation of the pyruvate carboxylation pathway. *Appl Microbiol Biotechnol* **2012**, *96(6)*, 1527-1537. doi: 10.1007/s00253-012-4192-z

143. Zhou, J.; Yin, X.; Madzak, C.; Du, G.; Chen, J. Enhanced α-ketoglutarate production in *Yarrowia lipolytica* WSH-Z06 by alteration of the acetyl-CoA metabolism. *J Biotechnol* **2012**, *161(3)*, 257-264. doi: 10.1016/j.jbiotec.2012.05.025

144. Walker, C.; Ryu, S.; Na, H. *et al*. Draft Genome Assemblies of Five Robust *Yarrowia lipolytica* Strains Exhibiting High Lipid Production, Pentose Sugar Utilization, and Sugar Alcohol Secretion from Undetoxified Lignocellulosic Biomass Hydrolysates. *Microbiol Resour Announc* **2018**, *7(12)*, e01040-18. doi:10.1128/MRA.01040-18

145. Kerscher, S.; Durstewitz, G.; Casaregola, S.; Gaillardin, C.; Brandt, U. The complete mitochondrial genome of *yarrowia lipolytica*. *Comp Funct Genomics* **2001**, *2(2)*, 80-90. doi: 10.1002/cfg.72

146. Blanchin-Roland, S.; Cordero Otero, R.R.; Gaillardin, C. Two upstream activation sequences control the expression of the *XPR2* gene in the yeast *Yarrowia lipolytica*. *Mol Cell Biol* **1994**, *14(1)*, 327-338. doi: 10.1128/mcb.14.1.327 1994

147. Madzak, C.; Mimmi, M.C.; Caminade, E. *et al*. Shifting the optimal pH of activity for a laccase from the fungus *Trametes versicolor* by structure-based mutagenesis. Protein Eng Des Sel **2006**, *9(2)*, 77-84. doi: 10.1093/protein/gzj004

148. Galli, C.; Gentili, P.; Jolivalt, C.; Madzak, C.; Vadalà, R. How is the reactivity of laccase affected by single-point mutations? Engineering laccase for improved activity towards sterically demanding substrates. *Appl Microbiol Biotechnol* **2011**, *91(1)*, 123-131. doi: 10.1007/s00253-011-3240-4

149. Bredeweg, E.L.; Pomraning, K.R.; Dai, Z.; Nielsen, J.; Kerkhoven, E.J.; Baker, S.E. A molecular genetic toolbox for *Yarrowia lipolytica*. *Biotechnol Biofuels* **2017**, *10*, 2. doi: 10.1186/s13068-016-0687-7

150. Wagner, J.M.; Williams, E.V.; Alper, H.S. Developing a piggyBac Transposon System and Compatible Selection Markers for Insertional Mutagenesis and Genome Engineering in *Yarrowia lipolytica*. *Biotechnol J* **2018**, *13(5)*, e1800022. doi: 10.1002/biot.201800022

151. Wang, J.; Ledesma-Amaro, R.; Wei, Y.; Ji, B.; Ji, X.J. Metabolic engineering for increased lipid accumulation in *Yarrowia lipolytica* - A Review. *Bioresour Technol* **2020**, *313*, 123707. doi: 10.1016/j.biortech.2020.123707

152. Dulermo, T.; Nicaud, J.M. Involvement of the G3P shuttle and β-oxidation pathway in the control of TAG synthesis and lipid accumulation in *Yarrowia lipolytica*. *Metab Eng* **2011**, *13(5)*, 482-491. doi: 10.1016/j.ymben.2011.05.002

153. Blazeck, J.; Hill, A.; Liu, L. *et al*. Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production. *Nat Commun* **2014**, *5*, 3131. doi: 10.1038/ncomms4131

154. Qiao, K.; Imam Abidi, S.H.; Liu, H. *et al*. Engineering lipid overproduction in the oleaginous yeast *Yarrowia lipolytica*. *Metab Eng* **2015**, *29*, 56-65. doi: 10.1016/j.ymben.2015.02.005

155. Lazar, Z.; Dulermo, T.; Neuvéglise, C.; Crutz-Le Coq, A.M.; Nicaud, J.M. Hexokinase--A limiting factor in lipid production from fructose in *Yarrowia lipolytica*. *Metab Eng* **2014**, *26*, 89-99. doi: 10.1016/j.ymben.2014.09.008

156. Lazar, Z.; Gamboa-Meléndez, H.; Le Coq, A.M.; Neuvéglise, C.; Nicaud, J.M. Awakening the endogenous Leloir pathway for efficient galactose utilization by *Yarrowia lipolytica*. *Biotechnol Biofuels* **2015**, *8*, 185. doi: 10.1186/s13068-015-0370-4

157. Liu, L.; Pan, A.; Spofford, C.; Zhou, N.; Alper, H.S. An evolutionary metabolic engineering approach for enhancing lipogenesis in *Yarrowia lipolytica*. *Metab Eng* **2015**, *29*, 36-45. doi: 10.1016/j.ymben.2015.02.003

158. Bordes, F.; Fudalej, F.; Dossat, V.; Nicaud, J.M.; Marty, A. A new recombinant protein expression system for high-throughput screening in the yeast *Yarrowia lipolytica*. *J Microbiol Methods* **2007**, *70(3)*, 493-502. doi: 10.1016/j.mimet.2007.06.008

159. Emond, S.; Montanier, C.; Nicaud, J.M. *et al*. New efficient recombinant expression system to engineer *Candida antarctica* lipase B. *Appl Environ Microbiol* **2010**, *76(8)*, 2684-2687. doi: 10.1128/AEM.03057-09

160. Leplat, C.; Nicaud, J.M.; Rossignol, T. Overexpression screen reveals transcription factors involved in lipid accumulation in *Yarrowia lipolytica*. *FEMS Yeast Res* **2018**, *18(5)*, 10.1093/femsyr/foy037. doi: 10.1093/femsyr/foy037

161. De Pourcq, K.; Tiels, P.; Van Hecke, A.; Geysens, S.; Vervecken, W.; Callewaert, N. Engineering *Yarrowia lipolytica* to produce glycoproteins homogeneously modified with the universal Man3GlcNAc2 N-glycan core. *PLoS One* 2012, 7(6), e39976. doi: 10.1371/journal.pone.0039976

162. Jost, B.; Holz, M.; Aurich, A.; Barth, G.; Bley, T.; Müller, R.A. The influence of oxygen limitation for the production of succinic acid with recombinant strains of *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **2015**, *99(4)*, 1675-1686. doi: 10.1007/s00253-014-6252-z

163. Yovkova V, Otto C, Aurich A, Mauersberger S, Barth G. 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* **2014**, *98(5)*, 2003-2013. doi: 10.1007/s00253-013-5369-9

164. Lazar, Z.; Neuvéglise, C.; Rossignol, T. *et al*. Characterization of hexose transporters in *Yarrowia lipolytica* reveals new groups of Sugar Porters involved in yeast growth. *Fungal Genet Biol* **2017**, *100*, 1-12. doi: 10.1016/j.fgb.2017.01.001

165. Wang, G.; Guo, L.; Liang, W.; Chi, Z.; Liu, L. Systematic analysis of the lysine acetylome reveals diverse functions of lysine acetylation in the oleaginous yeast *Yarrowia lipolytica*. *AMB Express* **2017**, *7(1)*, 94. doi: 10.1186/s13568-017-0393-2

166. Guo, H.; Wan, H.; Chen, H.; Fang, F.; Liu, S.; Zhou, J. Proteomic analysis of the response of α-ketoglutarate-producer *Yarrowia lipolytica* WSH-Z06 to environmental pH stimuli. *Appl Microbiol Biotechnol* **2016**, *100(20)*, 8829-8841. doi: 10.1007/s00253-016-7775-2

167. Larroude, M.; Rossignol, T.; Nicaud, J.M.; Ledesma-Amaro, R. Synthetic biology tools for engineering *Yarrowia lipolytica*. *Biotechnol Adv* **2018**, *36(8)*, 2150-2164. doi: 10.1016/j.biotechadv.2018.10.004

168. Darvishi, F.; Ariana, M.; Marella, E.R.; Borodina, I. Advances in synthetic biology of oleaginous yeast *Yarrowia lipolytica* for producing non-native chemicals. *Appl Microbiol Biotechnol* **2018**, *102(14)*, 5925-5938. doi: 10.1007/s00253-018-9099-x

169. Markham, K.A.; Alper, H.S. Synthetic Biology Expands the Industrial Potential of *Yarrowia lipolytica*. *Trends Biotechnol* **2018**, *36(10)*, 1085-1095. doi: 10.1016/j.tibtech.2018.05.004

170. Shi, T.Q.; Huang, H.; Kerkhoven, E.J.; Ji, X.J. Advancing metabolic engineering of *Yarrowia lipolytica* using the CRISPR/Cas system. *Appl Microbiol Biotechnol* **2018**, *102(22)*, 9541-9548. doi: 10.1007/s00253-018-9366-x

171. Ganesan, V.; Spagnuolo, M.; Agrawal, A.; Smith, S.; Gao, D.; Blenner, M. Advances and opportunities in gene editing and gene regulation technology for *Yarrowia lipolytica*. *Microb Cell Fact* **2019**, *18(1)*, 208. doi: 10.1186/s12934-019-1259-x

172. Vandermies, M.; Fickers, P. Bioreactor-Scale Strategies for the Production of Recombinant Protein in the Yeast *Yarrowia lipolytica*. *Microorganisms* **2019**, *7(2)*, 40. doi: 10.3390/microorganisms7020040

173. Chen, D.C.; Beckerich, J.M.; Gaillardin, C. One-step transformation of the dimorphic yeast *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **1997**, *48(2)*, 232-235. doi: 10.1007/s002530051043

174. Wang, J.H.; Hung, W.; Tsai, S.H. High efficiency transformation by electroporation of *Yarrowia lipolytica*. *J Microbiol* **2011**, *49(3)*, 469-472. doi: 10.1007/s12275-011-0433-6

175. Markham, K.A.; Vazquez, S.; Alper, H.S. High-efficiency transformation of *Yarrowia lipolytica* using electroporation. *FEMS Yeast Res* **2018**, *18(7)*, 10.1093/femsyr/foy081. doi: 10.1093/femsyr/foy081

176. Fournier, P.; Abbas, A.; Chasles, M. *et al*. Colocalization of centromeric and replicative functions on autonomously replicating sequences isolated from the yeast *Yarrowia lipolytica*. *Proc Natl Acad Sci U S A* **1993**, *90(11)*, 4912-4916. doi: 10.1073/pnas.90.11.4912

177. Vernis, L.; Abbas, A.; Chasles, M. *et al*. An origin of replication and a centromere are both needed to establish a replicative plasmid in the yeast *Yarrowia lipolytica*. *Mol Cell Biol* **1997**, *17(4)*, 1995-2004. doi: 10.1128/mcb.17.4.1995

178. Fickers, P.; Le Dall, M.T.; Gaillardin, C.; Thonart, P.; Nicaud, J.M. New disruption cassettes for rapid gene disruption and marker rescue in the yeast *Yarrowia lipolytica*. *J Microbiol Methods* **2003**, *55(3)*, 727-737. doi: 10.1016/j.mimet.2003.07.003

179. Wong, L.; Engel, J.; Jin, E.; Holdridge, B.; Xu, P. YaliBricks, a versatile genetic toolkit for streamlined and rapid pathway engineering in *Yarrowia lipolytica*. *Metab Eng Commun* **2017**, *5*, 68-77. doi: 10.1016/j.meteno.2017.09.001

180. Wong, L.; Holdridge, B.; Engel, J.; Xu, P. Genetic Tools for Streamlined and Accelerated Pathway Engineering in *Yarrowia lipolytica*. *Methods Mol Biol* **2019**, *1927*, 155-177. doi: 10.1007/978-1-4939-9142-6_11

181. Liu, L.; Otoupal, P.; Pan, A.; Alper, H.S. Increasing expression level and copy number of a *Yarrowia lipolytica* plasmid through regulated centromere function. *FEMS Yeast Res* **2014**, *14(7)*, 1124-1127. doi: 10.1111/1567-1364.12201

182. Juretzek, T.; Le Dall, M.; Mauersberger, S.; Gaillardin, C.; Barth, G.; Nicaud, J. Vectors for gene expression and amplification in the yeast *Yarrowia lipolytica*. *Yeast* **2001**, *18(2)*, 97-113. doi: 10.1002/1097-0061(20010130)18:2<97::AID-YEA652>3.0.CO;2-U

183. Larroude, M.; Park, Y.K.; Soudier, P.; Kubiak, M.; Nicaud, J.M.; Rossignol, T. A modular Golden Gate toolkit for *Yarrowia lipolytica* synthetic biology. *Microb Biotechnol* **2019**, *12(6)*, 1249-1259. doi: 10.1111/1751-7915.13427

184. Yue, L.; Chi, Z.; Wang, L. *et al*. Construction of a new plasmid for surface display on cells of *Yarrowia lipolytica*. *J Microbiol Methods* **2008**, *72(2)*, 116-123. doi: 10.1016/j.mimet.2007.11.011

185. Blazeck, J.; Liu, L.; Redden, H.; Alper, H. Tuning gene expression in *Yarrowia lipolytica* by a hybrid promoter approach. *Appl Environ Microbiol* **2011**, *77(22)*, 7905-7914. doi: 10.1128/AEM.05763-11

186. Celińska, E.; Ledesma-Amaro, R.; Larroude, M.; Rossignol, T.; Pauthenier, C.; Nicaud, J.M. Golden Gate Assembly system dedicated to complex pathway manipulation in *Yarrowia lipolytica*. *Microb Biotechnol* **2017**, *10(2)*, 450-455. doi: 10.1111/1751-7915.12605

187. Chuang, L.T.; Chen, D.C.; Nicaud, J.M.; Madzak, C.; Chen, Y.H.; Huang, Y.S. Co-expression of heterologous desaturase genes in *Yarrowia lipolytica*. *N Biotechnol* **2010**, *27(4)*, 277-282. doi: 10.1016/j.nbt.2010.02.006

188. Celińska, E.; Grajek, W. A novel multigene expression construct for modification of glycerol metabolism in *Yarrowia lipolytica*. *Microb Cell Fact* **2013**, *12*, 102. doi: 10.1186/1475-2859-12-102

189. Celińska, E.; Borkowska, M.; Korpys-Woźniak, P. *et al*. Optimization of *Yarrowia lipolytica*-based consolidated biocatalyst through synthetic biology approach: transcription units and signal peptides shuffling. *Appl Microbiol Biotechnol* **2020**, *104(13)*, 5845-5859. doi: 10.1007/s00253-020-10644-6

190. Holkenbrink, C.; Dam,M.I.; Kildegaard, K.R. *et al*. EasyCloneYALI: CRISPR/Cas9-Based Synthetic Toolbox for Engineering of the Yeast *Yarrowia lipolytica*. *Biotechnol J* **2018**, *13(9)*, e1700543. doi: 10.1002/biot.201700543

191. Celińska, E.; Ledesma-Amaro, R.; Larroude, M.; Rossignol, T.; Pauthenier, C.; Nicaud, J.M. Golden Gate Assembly system dedicated to complex pathway manipulation in *Yarrowia lipolytica*. *Microb Biotechnol* **2017**, *10(2)*, 450-455. doi: 10.1111/1751-7915.12605

192. Larroude, M.; Celinska, E.; Back, A.; Thomas, S.; Nicaud, J.M.; Ledesma-Amaro, R. A synthetic biology approach to transform *Yarrowia lipolytica* into a competitive biotechnological producer of β-carotene. *Biotechnol Bioeng* **2018**, *115(2)*, 464-472. doi: 10.1002/bit.26473

193. Gao, S.; Han, L.; Zhu, L. *et al*. One-step integration of multiple genes into the oleaginous yeast *Yarrowia lipolytica*. *Biotechnol Lett* **2014**, *36(12)*, 2523-2528. doi: 10.1007/s10529-014-1634-y

194. Gao, S.; Tong, Y.; Zhu, L. *et al*. Production of β-carotene by expressing a heterologous multifunctional carotene synthase in *Yarrowia lipolytica*. *Biotechnol Lett* **2017**, *39(6)*, 921-927. doi: 10.1007/s10529-017-2318-1

195. Liu, H.H.; Madzak, C.; Sun, M.L.; Ren, L.; Song, P. *et al*. Engineering *Yarrowia lipolytica* for arachidonic acid production through rapid assembly of metabolic pathway. *Biochem Eng J* **2017**, *119*, 52-58. doi: 10.1016/j.bej.2016.12.004

196. Guo, Z.P.; Borsenberger, V.; Croux, C. *et al*. An artificial chromosome ylAC enables efficient assembly of multiple genes in *Yarrowia lipolytica* for biomanufacturing. *Commun Biol* **2020**, *3(1)*, 199. doi: 10.1038/s42003-020-0936-y

197. Ogrydziak, D.; Demain, A.; Tannenbaum, S. Regulation of extracellular protease production in *Candida lipolytica*. *Biochim Biophys Acta* **1977**, *497*, 525–538. doi: 10.1016/0304-4165(77)90209-4

198. Müller, S.; Sandal, T.; Kamp-Hansen, P.; Dalbøge, H. 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* **1998**, *14(14)*, 1267-1283. doi: 10.1002/(SICI)1097-0061(1998100)14:14<1267::AID-YEA327>3.0.CO;2-2

199. Juretzek, T.; Wang, H.J.; Nicaud, J.M.; Mauersberger, S.; Barth, G. Comparison of promoters suitable for regulated overexpression of β-galactosidase in the alkane-utilizing yeast *Yarrowia lipolytica*. *Biotechnol Bioprocess Eng* **2000**, *5*, 320–326. doi: 10.1007/BF02942206

200. Trassaert, M.; Vandermies, M.; Carly, F. *et al*. New inducible promoter for gene expression and synthetic biology in *Yarrowia lipolytica*. *Microb Cell Fact* **2017**, *16(1)*, 141. doi: 10.1186/s12934-017-0755-0

201. Tai, M.; Stephanopoulos, G. Engineering the push and pull of lipid biosynthesis in oleaginous yeast *Yarrowia lipolytica* for biofuel production. *Metab Eng* **2013**, *15*, 1-9. doi: 10.1016/j.ymben.2012.08.007

202. Madzak, C.; Blanchin-Roland, S.; Cordero Otero, R.R.; Gaillardin, C. Functional analysis of upstream regulating regions from the *Yarrowia lipolytica XPR2* promoter. *Microbiology* **1999**, *145(1)*, 75-87. doi: 10.1099/13500872-145-1-75

203. Kopecný, D.; Pethe, C.; Sebela, M. *et al*. High-level expression and characterization of *Zea mays* cytokinin oxidase/dehydrogenase in *Yarrowia lipolytica*. *Biochimie* **2005**, *87(11)*, 1011-1022. doi: 10.1016/j.biochi.2005.04.006

204. Blazeck, J.; Reed, B.; Garg, R. *et al*. Generalizing a hybrid synthetic promoter approach in *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **2013**, *97(7)*, 3037-3052. doi: 10.1007/s00253-012-4421-5

205. Shabbir Hussain, M.; Wheeldon, I.; Blenner, M.A. A Strong Hybrid Fatty Acid Inducible Transcriptional Sensor Built From *Yarrowia lipolytica* Upstream Activating and Regulatory Sequences. *Biotechnol J* **2017**, *12(10)*, 10.1002/biot.201700248. doi: 10.1002/biot.201700248

206. Park, Y.K.; Korpys, P.; Kubiak, M. *et al*. Engineering the architecture of erythritol-inducible promoters for regulated and enhanced gene expression in *Yarrowia lipolytica*. *FEMS Yeast Res* **2019**, *19(1)*, 10.1093/femsyr/foy105. doi: 10.1093/femsyr/foy105

207. Dulermo, R.; Brunel, F.; Dulermo, T. *et al*. Using a vector pool containing variable-strength promoters to optimize protein production in *Yarrowia lipolytica*. *Microb Cell Fact* **2017**, *16(1)*, 31. doi: 10.1186/s12934-017-0647-3

208. Liu, R.; Liu, L.; Li, X.; Liu, D.; Yuan, Y. Engineering yeast artificial core promoter with designated base motifs. *Microb Cell Fact* **2020**, *19(1)*, 38. doi: 10.1186/s12934-020-01305-4

209. Curran, K.A.; Morse, N.J.; Markham, K.A.; Wagman, A.M.; Gupta, A., Alper, H.S. Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast. *ACS Synth Biol* **2015**, *4(7)*, 824-832. doi: 10.1021/sb5003357

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

211. Gasmi, N.; Fudalej, F.; Kallel, H.; Nicaud, J.M. A molecular approach to optimize hIFN α2b expression and secretion in *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **2011**, *89(1)*, 109-119. doi: 10.1007/s00253-010-2803-0

212. Celińska, E.; Borkowska, M.; Białas, W.; Korpys, P.; Nicaud, J.M. Robust signal peptides for protein secretion in Yarrowia lipolytica: identification and characterization of novel secretory tags. *Appl Microbiol Biotechnol* **2018**, *102(12)*, 5221-5233. doi: 10.1007/s00253-018-8966-9

213. Yuzbasheva, E.Y.; Yuzbashev, T.V.; Laptev, I.A.; Konstantinova, T.K.; Sineoky, S.P. Efficient cell surface display of Lip2 lipase using C-domains of glycosylphosphatidylinositol-anchored cell wall proteins of *Yarrowia lipolytica*. *Appl Microbiol Biotechnol* **2011**, *91(3)*, 645-654. doi: 10.1007/s00253-011-3265-8

214. Moon, H.Y.; Van, T.L.; Cheon, S.A.; Choo, J.; Kim, J.Y.; Kang, H.A. Cell-surface expression of *Aspergillus saitoi*-derived functional α-1,2-mannosidase on *Yarrowia lipolytica* for glycan remodeling. *J Microbiol* **2013**, *51(4)*, 506-514. doi: 10.1007/s12275-013-3344-x

215. Wang, J.M.; Wang, C.M.; Men, X. *et al*. Construction of arming *Yarrowia lipolytica* surface-displaying soybean seed coat peroxidase for use as whole-cell biocatalyst. *Enzyme Microb Technol* **2020**, *135*, 109498. doi: 10.1016/j.enzmictec.2019.109498

216. Yang, X.S.; Jiang, Z.B.; Song, H.T.; Jiang, S.J.; Madzak, C.; Ma, L.X. Cell-surface display of the active mannanase in *Yarrowia lipolytica* with a novel surface-display system. *Biotechnol Appl Biochem* **2009**, *54(3)*, 171-176. doi: 10.1042/BA20090222

217. Duquesne, S.; Bozonnet, S.; Borde, F.; Dumon, C.; Nicaud, J.M.; Marty, A. Construction of a highly active xylanase displaying oleaginous yeast: comparison of anchoring systems. *PLoS One* **2014**, *9(4)*, e95128. doi: 10.1371/journal.pone.0095128

218. Yuzbasheva, E.Y.; Yuzbashev, T.V.; Perkovskaya, N.I. *et al*. Cell surface display of *Yarrowia lipolytica* lipase Lip2p using the cell wall protein YlPir1p, its characterization, and application as a whole-cell biocatalyst. *Appl Biochem Biotechnol* **2015**, *175(8)*, 3888-3900. doi: 10.1007/s12010-015-1557-7

219. Haddouche, R.; Delessert, S.; Sabirova, J.; Neuvéglise, C.; Poirier, Y.; Nicaud, J.M. Roles of multiple acyl-CoA oxidases in the routing of carbon flow towards β-oxidation and polyhydroxyalkanoate biosynthesis in *Yarrowia lipolytica*. *FEMS Yeast Res* **2010**, *10(7)*, 917-927. doi: 10.1111/j.1567-1364.2010.00670.x

220. Kerkhoven, E.J.; Kim, Y.M.; Wei, S. *et al*. Leucine Biosynthesis Is Involved in Regulating High Lipid Accumulation in *Yarrowia lipolytica*. *mBio* **2017**, *8(3)*, e00857-17. doi: 10.1128/mBio.00857-17

221. Boeke, J.D.; Trueheart, J.; Natsoulis, G.; Fink, G.R. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol* **1987**, *154*, 164-175. doi: 10.1016/0076-6879(87)54076-9

222. Cheon, S.A.; Han, E.J.; Kang, H.A.; Ogrydziak, D.M.; Kim, J.Y. Isolation and characterization of the *TRP1* gene from the yeast *Yarrowia lipolytica* and multiple gene disruption using a TRP blaster. *Yeast* **2003**, *20(8)*, 677-685. doi: 10.1002/yea.987

223. Larroude, M.; Trabelsi, H.; Nicaud, J.M.; Rossignol, T. A set of *Yarrowia lipolytica* CRISPR/Cas9 vectors for exploiting wild-type strain diversity. *Biotechnol Lett* **2020**, *42(5)*, 773-785. doi: 10.1007/s10529-020-02805-4

224. Jang IS, Yu BJ, Jang JY, Jegal J, Lee JY. Improving the efficiency of homologous recombination by chemical and biological approaches in *Yarrowia lipolytica*. *PLoS One* **2018**, *13(3)*, e0194954. doi: 10.1371/journal.pone.0194954

225. Gaillardin, C.; Ribet, A.M. *LEU2* directed expression of beta-galactosidase activity and phleomycin resistance in *Yarrowia lipolytica*. *Curr Genet* **1987**, *11(5)*, 369-375. doi: 10.1007/BF00378179

226. Cordero Otero, R.; Gaillardin, C. Efficient selection of hygromycin-B-resistant *Yarrowia lipolytica* transformants. *Appl Microbiol Biotechnol* **1996**, *46(2)*, 143-148. doi: 10.1007/s002530050796

227. Lazar, Z.; Rossignol, T.; Verbeke, J.; Crutz-Le Coq, A.M.; Nicaud, J.M.; Robak, M. Optimized invertase expression and secretion cassette for improving *Yarrowia lipolytica* growth on sucrose for industrial applications. *J Ind Microbiol Biotechnol* **2013**, *40(11)*, 1273-1283. doi: 10.1007/s10295-013-1323-1

228. Vorachek-Warren, M.K.; McCusker, J.H. DsdA (D-serine deaminase): a new heterologous MX cassette for gene disruption and selection in *Saccharomyces cerevisiae*. *Yeast* **2004**, *21(2)*, 163-171. doi: 10.1002/yea.1074

229. Egermeier, M.; Sauer, M.; Marx, H. Golden Gate-based metabolic engineering strategy for wild-type strains of *Yarrowia lipolytica*. *FEMS Microbiol Lett* **2019**, *366(4)*, fnz022. doi: 10.1093/femsle/fnz022

230. Shi, T.Q.; Huang, H.; Kerkhoven, E.J.; Ji, X.J. Advancing metabolic engineering of *Yarrowia lipolytica* using the CRISPR/Cas system. *Appl Microbiol Biotechnol* **2018**, *102(22)*, 9541-9548. doi: 10.1007/s00253-018-9366-x

231. Doudna, J.A.; Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014**, *346(6213)*, 1258096. doi: 10.1126/science.1258096

232.   Schwartz, C.M.; Hussain, M.S.; Blenner, M.; Wheeldon, I. Synthetic RNA Polymerase III Promoters Facilitate High-Efficiency CRISPR-Cas9-Mediated Genome Editing in *Yarrowia lipolytica*. *ACS Synth Biol* **2016**, *5(4)*, 356-359. doi: 10.1021/acssynbio.5b00162

233.   Gao, S.; Tong, Y.; Wen, Z. *et al*. Multiplex gene editing of the *Yarrowia lipolytica* genome using the CRISPR-Cas9 system. *J Ind Microbiol Biotechnol* **2016**, *43(8)*, 1085-1093. doi: 10.1007/s10295-016-1789-8

234.   Schwartz, C.; Shabbir-Hussain, M.; Frogue, K.; Blenner, M.; Wheeldon, I. Standardized Markerless Gene Integration for Pathway Engineering in *Yarrowia lipolytica*. *ACS Synth Biol* **2017**, *6(3)*, 402-409. doi: 10.1021/acssynbio.6b00285

235.   Borsenberger, V.; Onésime, D.; Lestrade, D. *et al*. Multiple Parameters Drive the Efficiency of CRISPR/Cas9-Induced Gene Modifications in *Yarrowia lipolytica*. *J Mol Biol* **2018**, *430(21)*, 4293-4306. doi: 10.1016/j.jmb.2018.08.024

236.   Abdel-Mawgoud, A.M.; Stephanopoulos, G. Improving CRISPR/Cas9-mediated genome editing efficiency in *Yarrowia lipolytica* using direct tRNA-sgRNA fusions. *Metab Eng* **2020**, *62*, 106-115. doi: 10.1016/j.ymben.2020.07.008

237.   Yang, Z.; Edwards, H.; Xu, P. CRISPR-Cas12a/Cpf1-assisted precise, efficient and multiplexed genome-editing in *Yarrowia lipolytica*. *Metab Eng Commun* **2019**, *10:e00112*. doi: 10.1016/j.mec.2019.e00112

238.   Schwartz, C.; Cheng, J.F.; Evans, R. *et al*. Validating genome-wide CRISPR-Cas9 function improves screening in the oleaginous yeast *Yarrowia lipolytica*. *Metab Eng* **2019**, *55*, 102-110. doi: 10.1016/j.ymben.2019.06.007

239.   Scholze, H.; Boch, J. TAL effectors are remote controls for gene activation. *Curr Opin Microbiol* **2011**, *14(1)*, 47-53. doi: 10.1016/j.mib.2010.12.001

240.   Rigouin, C.; Gueroult, M.; Croux, C. *et al*. Production of Medium Chain Fatty Acids by *Yarrowia lipolytica*: Combining Molecular Design and TALEN to Engineer the Fatty Acid Synthase. *ACS Synth Biol* **2017**, *6(10)*, 1870-1879. doi: 10.1021/acssynbio.7b00034

241.   Rigouin, C.; Croux, C.; Dubois, G.; Daboussi, F.; Bordes, F. Genome Editing in *Y. lipolytica* Using TALENS. *Methods Mol Biol* **2021**, *2307*, 25-39. doi: 10.1007/978-1-0716-1414-3_2

242.   Shimatani, Z.; Kashojiya, S.; Takayama, M. *et al*. Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat Biotechnol* **2017**, *35(5)*, 441-443. doi: 10.1038/nbt.3833

243.   Bae, S.J.; Park, B.G.; Kim, B.G.; Hahn, J.S. Multiplex Gene Disruption by Targeted Base Editing of *Yarrowia lipolytica* Genome Using Cytidine Deaminase Combined with the CRISPR/Cas9 System. *Biotechnol J* **2020**, *15(1)*, e1900238. doi: 10.1002/biot.201900238

244.   Schwartz, C.; Frogue, K.; Ramesh, A.; Misa, J.; Wheeldon, I. CRISPRi repression of nonhomologous end-joining for enhanced genome engineering via homologous recombination in *Yarrowia lipolytica*. *Biotechnol Bioeng* **2017**, *114(12)*, 2896-2906. doi: 10.1002/bit.26404

245.   Zhang, J.L.; Peng, Y.Z.; Liu, D. *et al*. Gene repression via multiplex gRNA strategy in *Y. lipolytica*. *Microb Cell Fact* **2018**, *17(1)*, 62. doi: 10.1186/s12934-018-0909-8

246.   Schwartz, C.; Curtis, N.; Löbs, A.K.; Wheeldon, I. Multiplexed CRISPR Activation of Cryptic Sugar Metabolism Enables *Yarrowia Lipolytica* Growth on Cellobiose. *Biotechnol J* **2018**, *13(9)*, e1700584. doi: 10.1002/biot.201700584

247.   Misa, J.; Schwartz, C. CRISPR Interference and Activation to Modulate Transcription in *Yarrowia lipolytica*. *Methods Mol Biol* **2021**, *2307*, 95-109. doi: 10.1007/978-1-0716-1414-3_6

248.   Schwartz, C.; Wheeldon, I. CRISPR-Cas9-Mediated Genome Editing and Transcriptional Control in *Yarrowia lipolytica*. *Methods Mol Biol* **2018**, *1772*, 327-345. doi: 10.1007/978-1-4939-7795-6_18

249.   Patterson, K.; Yu, J.; Landberg, J. *et al*. Functional genomics for the oleaginous yeast *Yarrowia lipolytica*. *Metab Eng* **2018**, *48*, 184-196. doi: 10.1016/j.ymben.2018.05.008

250.   Loira, N.; Dulermo, T.; Nicaud, J.M.; Sherman, D.J. A genome-scale metabolic model of the lipid-accumulating yeast *Yarrowia lipolytica*. *BMC Syst Biol* **2012**, *6*, 35. doi: 10.1186/1752-0509-6-35

251.   Kerkhoven, E.J.; Pomraning, K.R.; Baker, S.E.; Nielsen, J. Regulation of amino-acid metabolism controls flux to lipid accumulation in *Yarrowia lipolytica*. *NPJ Syst Biol Appl* **2016**, *2*, 16005. doi: 10.1038/npjsba.2016.5

252.   Verbeke, J.; Beopoulos, A.; Nicaud, J.M. Efficient homologous recombination with short length flanking fragments in *Ku70* deficient *Yarrowia lipolytica* strains. *Biotechnol Lett* **2013**, *35(4)*, 571-576. doi: 10.1007/s10529-012-1107-0

253.   Kretzschmar, A.; Otto, C.; Holz, M.; Werner, S.; Hübner, L.; Barth, G. Increased homologous integration frequency in *Yarrowia lipolytica* strains defective in non-homologous end-joining. *Curr Genet* **2013**, *59(1-2)*, 63-72. doi: 10.1007/s00294-013-0389-7

254.   Ji, Q.; Mai, J.; Ding, Y.; Wei, Y.; Ledesma-Amaro, R.; Ji, X.J. Improving the homologous recombination efficiency of *Yarrowia lipolytica* by grafting heterologous component from *Saccharomyces cerevisiae*. *Metab Eng Commun* **2020**, *11*, e00152. doi: 10.1016/j.mec.2020.e00152

255.   Rosas-Quijano, R.; Gaillardin, C.; Ruiz-Herrera, J. Functional analysis of the *MATB* mating-type idiomorph of the dimorphic fungus *Yarrowia lipolytica*. *Curr Microbiol*. **2008**, *57(2)*, 115-120. doi: 10.1007/s00284-008-9162-4

256.   Han, C.; Kwon, H.; Park, G. *et al*. Enhanced mating-type switching and sexual hybridization in heterothallic yeast *Yarrowia lipolytica*. *FEMS Yeast Res* **2020**, *20(2)*, foaa011. doi: 10.1093/femsyr/foaa011

257.   McCarthy, C.G.P.; Fitzpatrick, D.A. Pangloss: A Tool for Pan-Genome Analysis of Microbial Eukaryotes. *Genes (Basel)* **2019**, *10(7)*, 521. doi: 10.3390/genes10070521

258.   Morin, N.; Cescut, J.; Beopoulos, A. *et al*. Transcriptomic analyses during the transition from biomass production to lipid accumulation in the oleaginous yeast *Yarrowia lipolytica*. *PLoS One* **2011**, *6(11)*, e27966. doi: 10.1371/journal.pone.0027966

259.   Sabra, W.; Bommareddy, R.R.; Maheshwari, G.; Papanikolaou, S.; Zeng, A.P. Substrates and oxygen dependent citric acid production by *Yarrowia lipolytica*: insights through transcriptome and fluxome analyses. *Microb Cell Fact* **2017**, *16(1)*, 78. doi: 10.1186/s12934-017-0690-0

260. Lazar, Z.; Liu, N.; Stephanopoulos, G. Holistic Approaches in Lipid Production by *Yarrowia lipolytica*. *Trends Biotechnol* **2018**, *36(11)*, 1157-1170. doi: 10.1016/j.tibtech.2018.06.007

261. Worland, A.M.; Czajka, J.J.; Xing, Y. *et al*. Analysis of *Yarrowia lipolytica* growth, catabolism, and terpenoid biosynthesis during utilization of lipid-derived feedstock. *Metab Eng Commun* **2020**, *11*, e00130. doi: 10.1016/j.mec.2020.e00130

262. Pan, P.; Hua, Q. Reconstruction and *in silico* analysis of metabolic network for an oleaginous yeast, *Yarrowia lipolytica*. *PLoS One* **2012**, *7(12)*, e51535. doi: 10.1371/journal.pone.0051535

263. Pomraning, K.R.; Wei, S.; Karagiosis, S.A. *et al*. Comprehensive Metabolomic, Lipidomic and Microscopic Profiling of *Yarrowia lipolytica* during Lipid Accumulation Identifies Targets for Increased Lipogenesis. *PLoS One* **2015**, *10(4)*, e0123188. doi: 10.1371/journal.pone.0123188

264. Trébulle, P.; Nicaud, J.M.; Leplat, C.; Elati, M. Inference and interrogation of a coregulatory network in the context of lipid accumulation in *Yarrowia lipolytica*. *NPJ Syst Biol Appl* **2017**, *3*, 21. doi: 10.1038/s41540-017-0024-1

265. Kim, M.; Park, B.G.; Kim, E.J.; Kim, J.; Kim, B.G. *In silico* identification of metabolic engineering strategies for improved lipid production in *Yarrowia lipolytica* by genome-scale metabolic modeling. *Biotechnol Biofuels* **2019**, *12*, 187. doi: 10.1186/s13068-019-1518-4

266. Wagner, J.M.; Liu, L.; Yuan, S.F.; Venkataraman, M.V.; Abate, A.R.; Alper, H.S. A comparative analysis of single cell and droplet-based FACS for improving production phenotypes: Riboflavin overproduction in *Yarrowia lipolytica*. *Metab Eng* **2018**, *47*, 346-356. doi: 10.1016/j.ymben.2018.04.015

267. Yang, X.; Wang, H.; Li, C.; Lin, C.S.K. Restoring of Glucose Metabolism of Engineered *Yarrowia lipolytica* for Succinic Acid Production via a Simple and Efficient Adaptive Evolution Strategy. *J Agric Food Chem* **2017**, *65(20)*, 4133-4139. doi: 10.1021/acs.jafc.7b00519

268. Li, J.; Zhu, K.; Miao, L. *et al*. Simultaneous Improvement of Limonene Production and Tolerance in *Yarrowia lipolytica* through Tolerance Engineering and Evolutionary Engineering. *ACS Synth Biol* **2021**, *10(4)*, 884-896. doi: 10.1021/acssynbio.1c00052

269. Qiu, X.; Gu, Y.; Du, G.; Zhang, J.; Xu, P.; Li, J. Conferring thermotolerant phenotype to wild type *Yarrowia lipolytica* improves cell growth and erythritol production. *Biotechnol Bioeng* **2021**, 10.1002/bit.27835. doi: 10.1002/bit.27835

270. Zhou, L.; Xu, Z.; Wen, Z. *et al*. Combined adaptive evolution and transcriptomic profiles reveal aromatic aldehydes tolerance mechanisms in *Yarrowia lipolytica*. *Bioresour Technol* **2021**, *329*, 124910. doi: 10.1016/j.biortech.2021.124910

271. Wang, Z.; Zhou, L.; Lu, M. *et al*. Adaptive laboratory evolution of *Yarrowia lipolytica* improves ferulic acid tolerance. *Appl Microbiol Biotechnol* **2021**, *105(4)*, 1745-1758. doi: 10.1007/s00253-021-11130-3

272. Xie, D. Integrating Cellular and Bioprocess Engineering in the Non-Conventional Yeast *Yarrowia lipolytica* for Biodiesel Production: A Review. *Front Bioeng Biotechnol* **2017**, *5*, 65. doi: 10.3389/fbioe.2017.00065

273. Braun, A.; Geier, M.; Bühler, B.; Schmid, A.; Mauersberger, S.; Glieder, A. Steroid biotransformations in biphasic systems with *Yarrowia lipolytica* expressing human liver cytochrome P450 genes. *Microb Cell Fact* 2012, 11, 106. doi: 10.1186/1475-2859-11-106

274. Kavšček, M.; Bhutada, G.; Madl, T.; Natter, K. Optimization of lipid production with a genome-scale model of *Yarrowia lipolytica*. *BMC Syst Biol* **2015**, *9*, 72. doi: 10.1186/s12918-015-0217-4

275. Gurme, S.T.; Surwase, S.N.; Patil, S.A.; Jadhav, J.P. Evaluation of Various Factors Affecting Bioconversion of l-Tyrosine to l-DOPA by Yeast *Yarrowia lipolytica*-NCIM 3450 Using Response Surface Methodology. *Nat Prod Bioprospect* **2014**, *4(3)*, 141-147. doi: 10.1007/s13659-014-0017-3

276. Rywińska, A.; Marcinkiewicz, M.; Cibis, E.; Rymowicz, W. Optimization of medium composition for erythritol production from glycerol by *Yarrowia lipolytica* using response surface methodology. *Prep Biochem Biotechnol* **2015**, *45(6)*, 515-529. doi: 10.1080/10826068.2014.940966

277. Yellapu, S.K.; Bezawada, J.; Kaur, R.; Kuttiraja, M.; Tyagi, R.D. Detergent assisted lipid extraction from wet yeast biomass for biodiesel: A response surface methodology approach. *Bioresour Technol* **2016**, *218*, 667-673. doi: 10.1016/j.biortech.2016.07.011

278. Darvishi, F.; Moradi, M.; Madzak, C.; Jolivalt, C. Production of Laccase by Recombinant Yarrowia lipolytica from Molasses: Bioprocess Development Using Statistical Modeling and Increase Productivity in Shake-Flask and Bioreactor Cultures. *Appl Biochem Biotechnol* **2017**, *181(3)*, 1228-1239. doi: 10.1007/s12010-016-2280-8

279. Celińska, E.; Nicaud, J.M.; Białas, W. Hydrolytic secretome engineering in *Yarrowia lipolytica* for consolidated bioprocessing on polysaccharide resources: review on starch, cellulose, xylan, and inulin. *Appl Microbiol Biotechnol* **2021**, *105(3)*, 975-989. doi: 10.1007/s00253-021-11097-1

280. Berg, P.; Baltimore, D.; Brenner, S.; Roblin, R.O.; Singer, M.F. Summary statement of the Asilomar conference on recombinant DNA molecules. *Proc Natl Acad Sci USA* **1975**, *72(6)*, 1981–1984. doi: 10.1073/pnas.72.6.1981

281. Gaskell, G.; Bauer, M.W.; Durant, J.; Allum, N.C. Worlds apart? The reception of genetically modified foods in Europe and the U.S. *Science* **1999**, *285(5426)*, 384–387. doi: 10.1126/science.285.5426.384

282. A CRISPR definition of genetic modification. *Nat Plants* **2018**, *4(5)*, 233. doi: 10.1038/s41477-018-0158-1

283. Ruffell, D. The EU Court of Justice extends the GMO Directive to gene-edited organisms. *FEBS Lett* **2018**, *592(22)*, 3653-3657. doi: 10.1002/1873-3468.13293

284. Tagliabue, G. The nonsensical GMO pseudo-category and a precautionary rabbit hole. *Nat Biotechnol* **2015**, *33(9)*, 907-908. doi: 10.1038/nbt.3333

285. Tagliabue, G. The meaningless pseudo-category of "GMOs": The trouble with the "new techniques" for genetically modifying crops demonstrates the illogical process-based definition of GMOs in EU regulation. EMBO Rep **2016**, *17(1)*, 10-13. doi: 10.15252/embr.201541385

286. Qiu, X.; Xu, P.; Zhao, X.; Du, G.; Zhang, J.; Li, J. Combining genetically-encoded biosensors with high throughput strain screening to maximize erythritol production in *Yarrowia lipolytica*. *Metab Eng* **2020**, *60*, 66-76. doi: 10.1016/j.ymben.2020.03.006