Molecular genetic tools for manipulation of the oleaginous yeast *Rhodotorula toruloides*

Submitted by Alexander Michael Bedford Johns to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Biological Sciences in December, 2016

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……

Acknowledgements

First I would like to thank my supervisor Dr Steve Aves for his patience and guidance during this project. I would also like to thank my second supervisor Dr John Love his input.

I would like to thank the Biotechnology and Biological Sciences Research Council (BBSRC) for funding my time in Exeter.

For teaching me so much of what was required to undertake a PhD, I would like to thank Dr Elizabeth Bilsland and Dr Balint Kintses.

I would like to thank my fellow SWDTP class of 2016 Exeter cohort Andrea Gori and Odette Wills for their friendship and support.

For their help during my first rotation project I would like to thank Prof. Nicholas Smirnoff, Dr Choon Lim, Dr Chloe Singleton and Dr Hannah Florence.

For their help, friendship and input during this project I would like to thank my colleagues in the Mezzanine lab including, but not limited to, Dr Anja Nenninger, Dr Christine Sambles, Dagmara Kolak, Dr Dayananda Chandrappa, Dr Debbie Salmon, George Kirke, Dr Richard Tennant, Ross Kent, Dr Thomas Lux, Sebastian Bowyer and Dr Steve Brown. I would also like to thank Dr Martin Schuster, Dr Sreedhar Kilaru and Dr Yaadwinder Sidhu for their help and input during this project.

I would like to thank my teammates during Biotechnology Yes: Hannah Daniels, Jamie Gilman, Katie Jones and Peggy Dousseaud.

Finally I would like to thank my Mum, Dad and Juliette for everything they have done to help and support me over the last four years and beyond.

Abstract

Rhodotorula (*Rhodosporidium*) *toruloides* is an oleaginous basidiomycete yeast with great biotechnological potential. Capable of accumulating lipid up to 76 % of its dry biomass and well suited to the metabolism of lignocellulosic hydrolysate, it is a good candidate for production of advanced biofuels as well as a host of other potential roles in industry. However, molecular genetic tools for manipulation of this yeast are lacking and its high genomic GC content can make routine cloning difficult. *Agrobacterium tumefaciens*-mediated transformation of *R. toruloides* CBS 14 was demonstrated, and plasmid vectors were developed for transformation of *R. toruloides*, including elements for *Saccharomyces cerevisiae* in-yeast assembly. In-yeast assembly is robust to the manipulation of GC-rich DNA and of large plasmids. Using these vectors and an EGFP reporter, a screen to identify inducible promoters was performed, and promoters from the genes *NAR1*, *ICL1*, *CTR3*, and *MET16* identified. These promoters have independent induction/repression conditions and different levels and rates of induction. Minimal inducible promoters were determined, which are as small as 200 bp. As well as showing tight regulation of the EGFP marker, the *NAR1* promoter was able to drive conditional rescue of a *leu2* mutant strain. In parallel, as a proof of principle for production of advanced biofuels, hydrocarbon biosynthesis pathways were expressed in *R. toruloides* and analysed by GC-MS. After co-expression of *Synechococcus elongatus* fatty acyl-ACP reductase and fatty aldehyde decarbonylase, and *E. coli* ferredoxin and ferredoxin reductase, production of the alkane heptadecane was observed. To increase the availability of free fatty acids (FFA) for production of hydrocarbons by other pathways, *Thermomyces lanuginosus* lipase 2 was expressed, resulting in a 1.3-fold increase in the concentration of FFAs.

Contents

List of tables

List of figures

List of abbreviations

ACL	ATP citrate lyase
ARS	Autonomously replicating sequence
ATMT	*Agrobacterium tumefaciens*-mediated transformation
BCS	Bathocuproinedisulfonic acid
DSB	Double strand break
EGFP	Enhanced green fluorescent protein
EIC	Extracted ion chromatogram
GC-FID	Gas chromatography – flame ionising detection
GC-MS	Gas chromatography – mass spectrometry
GFP	Green fluorescent protein
FAEE	Fatty acid ethyl ester
FAME	Fatty acid methyl ester
FAS	Fatty acid synthase
FFA	Free fatty acid
MCS	Multiple cloning site
OAA	Oxaloacetate
ORI	Origin of replication
PAH	Phenylalanine hydroxylase
PAL	Phenylalanine ammonia lyase
PCR	Polymerase chain reaction
PKU	Phenylketonuria
SCO	Single cell oil
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis

TAG	Triacylglycerol
TAIL-PCR	Thermal asymmetric interlaced-polymerase chain reaction
TBAH	Tetrabutylammonium hydroxide
UTR	Untranslated region
YFP	Yellow fluorescent protein

1 Introduction

1.1 Microbial biofuels

Climate change and finite oil reserves mean our current global reliance on fossil fuels is unsustainable. Combined with uncertainty over oil prices and calls for energy independence in the United States of America (United States Congress, 2007) mean development of sustainable alternatives is of paramount importance. While solar, wind, nuclear and other technologies can displace fossil fuels for electricity generation, liquid fuels will still be required for the transport sector (de Jong *et al.*, 2012, Lee *et al.*, 2015).

Biofuels are a promising alternative to petroleum derived fuels, reflected in governmental energy policies around the world; for example the US energy policy act of 2005 offers tax incentives to promote use of biofuels (United States Congress, 2005) and the European 2020 climate and energy package stipulates that by 2020 at least 10 % of transport fuels sold should be biofuels (European Commission, 2009).

First generation biofuels such as bioethanol produced from sugarcane or biodiesel produced by transesterification of plant oils are flawed. Their production is problematic as large areas of land are required for cultivation of feedstocks; this land use requirement places biofuels in competition with production of food, or in cases where previously uncultivated land is used can cause a net increase in carbon emissions relative to fossil fuels (Searchinger *et al.*, 2008). Furthermore, incompatibilities with existing vehicles and infrastructure limit their uptake (Fazal *et al.*, 2011). Therefore, despite the

promise offered by biofuels, the technology must be improved (Naik *et al.*, 2010). Different fuels will require different solutions for their replacement. One potential technology is the use of single cell oils (SCO) and derivatives thereof for the replacement of petrodiesel.

Autotrophic microorganisms have been the subject of research for SCO production while reducing the land use burden associated with first generation biofuels. The microalga *Botryococcus braunii* has attracted special interest as it is inherently capable of accumulating hydrocarbons and lipids which could be used for production of fuels (Dayananda *et al.*, 2007); however it has a low growth rate, with a highest reported doubling time of 1.4 days and can only achieve low culture densities (Li *et al.*, 2008, Yoshimura *et al.*, 2013). As a result the volumetric productivity, and therefore yields are low. This is a common problem with autotrophic microorganisms and, although appealing, their cultivation is not economically viable (Liao *et al.*, 2016).

Heterotrophic microorganisms suffer the disadvantage that they require provision of a food source; however they can be cultivated faster and to much higher densities than is feasible in systems where light is required for photosynthesis, to the extent that cultivation of potentially autotrophic microalgae can be more cost-effectively grown under heterotrophic conditions (Perez-Garcia *et al.*, 2011). If heterotrophic microorganisms were grown using a sustainable feedstock they could potentially be a sustainable, economically viable alternative to petrodiesel or first generation biodiesel.

1.1.1 Requirements of heterotrophic microorganisms for biofuel production

Due to the relatively low value of fuels, for microbial production to be viable the organism(s) used must be able to metabolise a low cost, sustainably sourced feedstock, yielding a higher value product, by means of an efficient, low cost process. Organisms capable of this do not exist in the wild, however through metabolic engineering or 'synthetic biology' it should be possible to engineer a microorganism, or a consortium of microorganisms to perform this role. It is beyond the scope of this work to fully review all potential microorganisms, however I shall summarise the key attributes of a good host for biofuel production before describing why *Rhodotorula* (*Rhodosporidium*) *toruloides* is a strong candidate.

Sugar or corn are commonly used as a feedstocks for microbial fermentation. Although cost effective for production of high-value compounds, in the absence of subsidies this is prohibitively expensive for production of low-value commodities such as fuel. For example in June 2016 the wholesale price of sugar was 0.375 USD kg^{-1} whereas the wholesale price of diesel was 0.307 USD kg^{-1} meaning that even in the case of perfect conversion of the carbon in the sugar to diesel, the feedstock used would cost more than the fuel produced. Compounding this is the environmentally unsustainable nature of sugar cultivation and competition for land use with food production.

Lignocellulosic biomass is the most abundant raw material on earth for production of fuels and is currently the most promising feedstock for sustainable, cost-effective production of microbial biofuels (Steen *et al.*, 2010,

Isikgor & Becer, 2015). This is the woody material from plants and can be produced by fast growing energy crops such as *Miscanthus* (Lewandowski *et al.*, 2000) or from agricultural or other waste streams (Isikgor & Becer, 2015). This material is composed of polysaccharides (cellulose and hemicelluloses), which can be hydrolysed to liberate five- and six-carbon sugars, and the aromatic polymer lignin. By virtue of its role as a structural material it is tough and recalcitrant to degradation. While there are fungi capable of degrading lignocellulosic biomass, biological degradation of this material is slow (Couturier *et al.*, 2012, Kang *et al.*, 2014). Therefore, to facilitate use of this material as a feedstock for microbial fermentation, it must be pretreated to disrupt the lignin layer that shields polysaccharides and to disrupt the crystalline structure of the cellulose facilitating hydrolysis (Kumar *et al.*, 2009). Several different methodologies for pretreatment of this material have been developed, however cost-effective processes liberate compounds inhibitory to the growth of microorganisms, including organic acids, phenolics and aldehydes (Kumar *et al.*, 2009). Therefore to use lignocellulose biomass as feedstock, microorganisms must be able to both metabolise the five- and six-carbon sugars released by the hydrolysis of the polysaccharides and be resistant to the inhibitory compounds liberated during pretreatment.

Biodiesel is produced by transesterification of triacylglycerol (TAG) with simple alcohols, commonly methanol or ethanol yielding fatty acid methyl esters (FAMEs) or fatty acid ethyl esters (FAEEs) respectively. Therefore to use microbes for production of this biodiesel, microbes must accumulate lipids. While biodiesel produced in this manner has the potential to be environmentally sustainable (Liao *et al.*, 2016), these fatty esters are

incompatible with certain materials found in some engines including rubber and copper-based alloys, at low temperatures they suffer from gelling limiting their usefulness in cooler climates, and transesterification can result in the presence of corrosive or hygroscopic contaminants (Fazal *et al.*, 2011). *In vivo* transesterification could reduce costs and potential contaminants, but would not overcome the fundamental problems associated with biodiesel (Lee *et al.*, 2015). As a result currently marketed biodiesel is mixed with petrodiesel, with B7 (7 % biodiesel) being the most common blend sold in Europe (Kampman *et al.*, 2013). During the last decade biological pathways for production of hydrocarbons from fatty acids have been described allowing for production of 'drop-in' petroleum replacement fuels. These pathways will be discussed in chapter 5, however the ideal microorganisms for production of hydrocarbon fuels must be able to accumulate lipid, and be tractable enough that pathways for hydrocarbon biosynthesis can be expressed.

Producing a microorganism capable of metabolising lignocellulosic biomass hydrolysate to rapidly and efficiently produce drop-in petroleum replacement fuel will require genetic engineering. Commonly used model organisms *Escherichia coli* and *Saccharomyces cerevisiae* have been used to develop technologies required for biofuel production (Steen *et al.*, 2010, Howard *et al.*, 2013, Zhou *et al.*, 2014, Zhou *et al.*, 2016), and due to the pre-existing knowledge surrounding these organisms and their ease of manipulation they make appealing platforms for biofuel production. For example, in *S. cerevisiae* research has been undertaken to increase its lipid yield and use it as a chassis for hydrocarbon biosynthesis. By deregulation of lipid biosynthesis through overexpression of a truncated diacylglycerol acyltransferase (*DGA1*)

in a *dga1*Δ *snf2*Δ background, a lipid content of 45 % was achieved, 90 % of which was TAGs (Kamisaka *et al.*, 2013). However, at this concentration the accumulated lipid was toxic to the host and further increases in lipid accumulation would require a much better understanding of how metabolism is regulated in *S. cerevisiae*. Several oleaginous organisms are natively capable of accumulating lipids to higher concentrations and are suited to metabolism of lignocellulosic biomass (Table 1.1); therefore a preferable strategy could be to use an unconventional, oleaginous organism as a chassis for microbial biofuel production.

1.1.2 Oleaginous microorganisms

Under conditions of excess carbon relative to other nutrients, most commonly nitrogen, many organisms accumulate lipids suitable for use in production of fuels. A minority of bacteria (most store excess carbon as polyhydroxyalkanoates which are not suitable for biofuel production (Sudesh *et al.*, 2000)) and some heterotrophic microalgae can achieve high culture densities and lipid yields, and therefore may constitute a viable platform for biofuel production. However the majority of microorganisms with properties suited to production of biofuels are oleaginous yeasts (Garay *et al.*, 2014). Examples of lipid yields achieved in oleaginous microorganisms, with various economically relevant (and otherwise) carbon sources are listed in Table 1.1. Also included is *B. braunii* as an example of yields achievable in an autotrophic organism.

Table 1.1: Oleaginous microorganisms with potential for biofuel production.

Organism	Carbon source	Culture conditions	Time (h)	Biomass (g L^{-1})	Lipid yield (g L^{-1})	% Lipid	Lipid coefficient	Productivity (g L^{-1} h^{-1})	Reference
Yeasts and fungi									
Rhodotorula toruloides	Glucose	Shake flask	120	18	13.9	76	0.22		Li *et al.*, 2006
	Lignocellulosic hydrolysate	Fed batch	96	54	31	59	0.29	0.6	Fei *et al.*, 2016
Lipomyces starkeyi	Glucose	Shake flask	120	28	17.1	57	0.18		Gong *et al.*, 2012
	Cellobiose, xylose	Shake flask	108	25.9	13.4	52	0.2		Gong *et al.*, 2012
Yarowia lypotica	Sugarcane bagasse hydrolysate	Shake flask	96	19.4	11.4	58			Tsigie *et al.*, 2011
	Glycerol	Continuous culture		60	24.2	40	0.1	0.43	Rakicka *et al.*, 2015
Trichosporon dermatis	Corncob enzymatic hydrolysate	Shake flask	168	24.4	9.8	40	0.16		Huang *et al.*, 2012
Cryptococcus curvatus	Glycerol	Fed batch	134	58.9	24.7	43	0.54	0.42	Thiru *et al.*, 2011
Algae									
Botrycoccus braunii	CO_2	Shake flask	1008	2	0.9[1]	46[1]			Dayananda *et al.*, 2007
Chlorella protothecoids	Glucose	Shake flask	144	3.9	2.14	55			Xu *et al.*, 2006
	Corn powder hydrolysate	Shake flask	144	3.7	2.06	55			Xu *et al.*, 2006
Bacteria									
Rhodococcus opacus	Dairy waste water	Batch	48	3.5	1.89	51			Kumar *et al*., 2015
Gordonia sp.	Orange waste	Shake flask	96	1.88	1.5	80			Gouda *et al.*, 2008

[1] Total organic fraction, includes lipids and hydrocarbons

Table 1.1 only summarises a small fraction of published studies on oleaginous microorganisms. However, not all organisms have been tested for growth on all potential feedstocks and it is difficult to directly compare different studies as they have often been performed under different conditions which can impact observed results, for example different feeding strategies can give very different yields (Fei *et al.*, 2016). Also, many of these organisms could be 'improved' with directed evolution or genetic engineering, meaning these studies may not account for the full potential of each organism (Liu *et al.*, 2015a, Zhang *et al.*, 2016). Furthermore there are other biological and practical factors which may impact choice of host, for example the ability to grow at low pH or high temperature (Liao *et al.*, 2016), ease of manipulation and pre-existing knowledge may be assets for some organisms, whereas intellectual property or potential pathogenicity may obstruct their use.

Yarrowia lipolytica has become a model organism for oleaginous yeast (Nicaud *et al.*, 2014), with a well annotated genome (Dujon *et al.*, 2004), developed molecular genetics (Nicaud *et al.*, 2014) and metabolic models (Pan & Hua, 2012). However this yeast is unable to metabolise some sugars liberated by hydrolysis of lignocellulosic biomass (Ryu *et al.*, 2015) and there are other yeasts that can accumulate greater lipid yields.

1.1.3 *Rhodotorula toruloides* for production of biodiesel

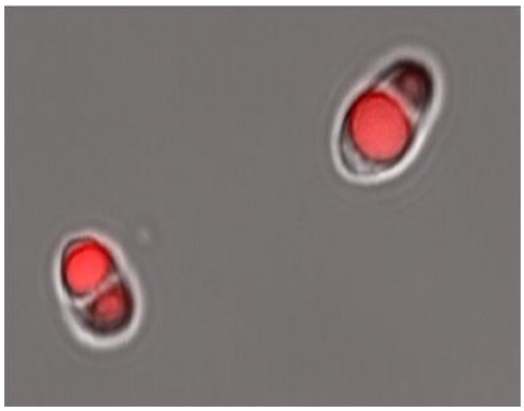


Figure 1.1 *R. toruloides* grown under lipid accumulating conditions stained with Nile red showing lipid droplets in red.

R. toruloides is a non-pathogenic (cf. *Trichosporon dermatis*) oleaginous yeast and is one of the best candidates for production of SCO-derived fuels. *R. toruloides* can metabolise the five- and six-carbon sugars liberated by degradation of lignocellulosic biomass, and has been demonstrated to achieve a high lipid yield when using lignocellulosic enzymatic hydrolysate as a carbon source (Table 1.1) (Fei *et al.*, 2016). Although *R. toruloides* is sensitive to toxic compounds liberated during acid hydrolysis of lignocellulosic biomass, directed evolution has been used to generate highly resistant stains able to give high lipid yields even in the presence of these compounds (Qi *et al.*, 2014). Furthermore, *R. toruloides* can metabolise other low cost, economically relevant feedstocks including acetic acid (Huang *et al.*, 2016), glycerol (Xu *et al.*, 2012) and inulin (Wang *et al.*, 2014).

When grown in media replete with carbon but limited for inorganic nitrogen, phosphorous or sulfur (Wu *et al.*, 2010, Wu *et al.*, 2011) or with organic compounds as the sole nitrogen source (Evans & Ratledge, 1984b) *R.*

toruloides accumulates lipid, mostly as TAGs (Figure 1.1). Through manipulation of the growth conditions a lipid content 76 % w/w can be achieved (Li *et al.*, 2006). Furthermore, *R. toruloides* is relatively fast growing with a doubling time of 160 minutes at 27 °C in rich media (Abe *et al.*, 1984) and can be cultivated to high densities, which, in combination with the high lipid content gives a high lipid yield (Table 1.1) (Li *et al.*, 2007, Fei *et al.*, 2016). The fatty acid composition of lipid produced by *R. toruloides* is approximately 50 % oleic acid (C18:1), ~20 % palmitic acid (C16:0), ~10 % stearic acid (C18:0) ~10 % linoleic acid (C18:2), with lower amounts of other lipids including linolenate (C18:3), palmitoleic (C16:1), tetracosanoic (C24:0) and myristic (C14:0) acids (Li *et al.*, 2007, Wiebe *et al.*, 2012). Although no one lipid blend will be optimal for every market, this high percentage of monounsaturated fatty acids and low proportion of polyunsaturates is a good mixture for production of biodiesel (Refaat, 2009).

Compared with other, more commonly used organisms, the intellectual property landscape surrounding *R. toruloides* is relatively unrestrictive. At the time of writing (June 2016) the World Intellectual Property Organisation lists 29 patents directly relating to *Rhodosporidium* or *Rhodotorula*, nine of which pertain to lipid / biofuel production or manipulation techniques which could potentially impact biodiesel production in *R. toruloides* (WO/2008/121701, WO/2016/039685, WO/2012/169969, WO/2014/142747, WO/2014/198988, WO/2016/016805, WO/2016/108185, WO/2015/127421, WO/2016/017183). In comparison there are 45 patents, 35 of which pertain to lipid / biofuel production or manipulation techniques relating to *Y. lipolytica*.

While *R. toruloides* is a promising system for production of biodiesel, in common with other oleaginous yeasts it could be 'improved' by genetic engineering. As previously discussed, expression of hydrocarbon biosynthesis pathways and increasing resistance to the toxic compounds found in lignocellulosic biomass hydrolysate would improve the utility of this organism as a system for production of fuels. Another avenue for improvement, in common with other oleaginous yeasts would be to engineer cells for easier lipid recovery. Lipid accumulates as lipid bodies within the cytoplasm, therefore cells must be lysed to extract lipid. *R. toruloides* can be cultivated to high density with a high proportion of lipid, therefore high yields are achieved in batch culture (Li *et al.*, 2007). However, were it possible to recover lipid from a continuous culture without killing and lysing cells, costs could be reduced as the tough cell wall makes processing of biomass and lipid recovery difficult and therefore expensive (Jacob, 1992), and decoupling cell growth and lipid production could allow an increased amount of carbon from the feedstock to be used for lipid production, increasing the lipid coefficient (Doshi *et al.*, 2013). As a result, while *R. toruloides* is a good chassis for biofuel production, through engineering it could be made better.

Relative to model organisms or well established biotechnological yeasts or fungi, the molecular genetic tools to facilitate engineering of *R. toruloides* are lacking. Therefore the first challenge for development of *R*. *toruloides* as a system for production or biofuels, and the overall aim of this project is development of tools and protocols for manipulation of this yeast.

1.2 *R. toruloides* for biotechnology

1.2.1 Phylogeny of *R. toruloides* and related strains

R. toruloides is a member of the Pucciniomycotina subphylum of the Basydiomycota. Evolutionarily, this puts it distant from other common laboratory or industrial yeasts such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Komagatealla* (*Pichia*) *pastoris*, *Aspergillus nidulans* or *Penicillium* moulds, all of which are ascomycetes (Figure 1.2). Similarly, *Y. lipolytica*, probably the best studied of the oleaginous yeasts, is also an ascomycete. From a molecular genetic perspective the best studied basidiomycete would most likely be the human pathogen *Cryptococcus neoformans* or the economically important plant pathogen *Ustilago maydis*; however even these organisms are only distantly related to *R. toruloides*, belonging to the Agaricomycotina and Ustilaginomycotina respectively, with very different lifestyles and ecological niches. Therefore, there is no single closely related, physiologically similar, well-characterised system on which to base study of *R. toruloides*.

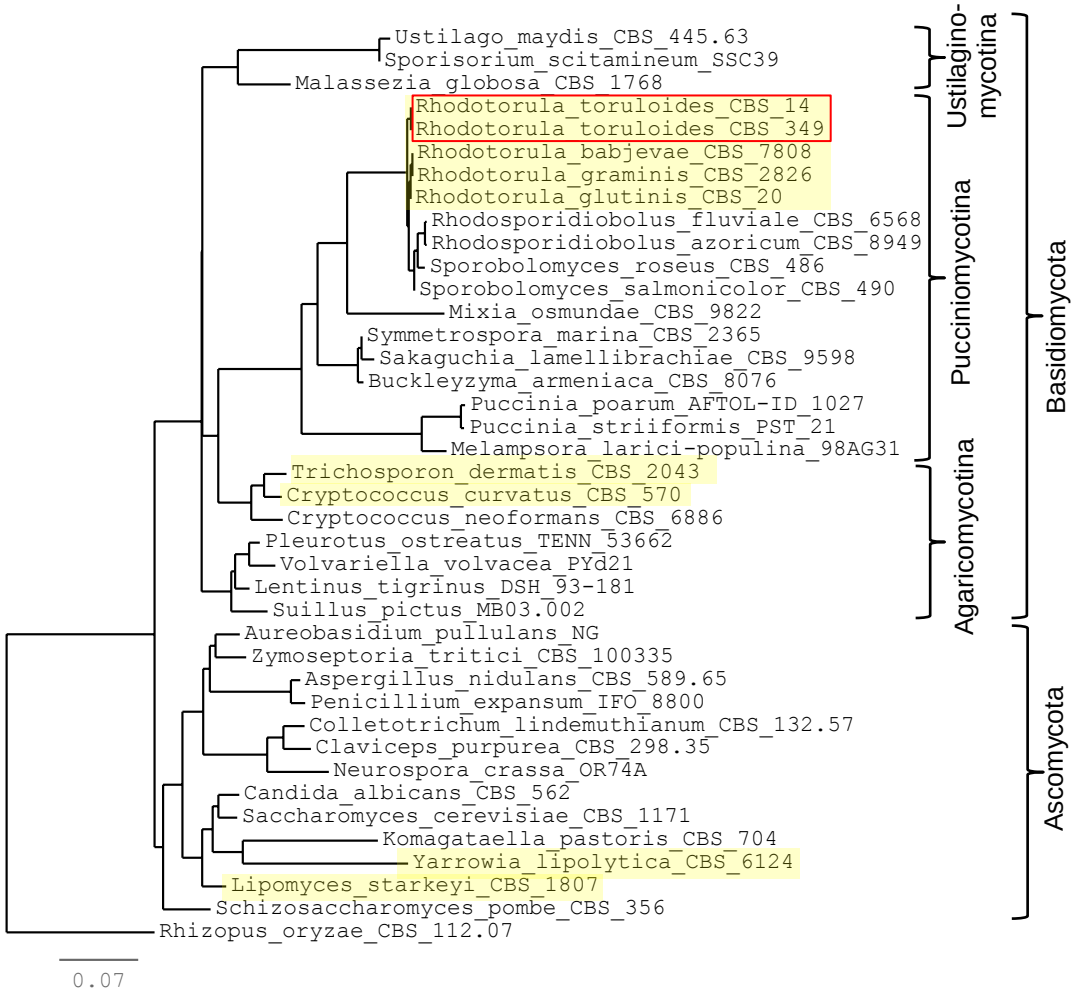


Figure 1.2. 18S ribosomal RNA cladogram showing the relationship between *R. toruloides* (red box) and other members of the fungal subkingdom Dikarya. *Rhizopus oryzae* (phylum: Zygomycota) is included as an outgroup in order to root the tree. Oleaginous yeasts previously studied for their potential in biofuel production are highlighted in yellow. Scale bar indicates number of substitutions per site for a given branch length. Tree constructed using the phylogeny.fr platform in one click mode (Dereeper *et al.*, 2008).

Rhodotorula is a cosmopolitan genius, with examples being isolated in a wide variety of habitats including decaying conifer wood in Sweden (*R. toruloides* CBS 14), Japanese air (*R. toruloides* CBS 349) and the gut of a porpoise in the Bahamas (*R. toruloides* CBS 5490). However the decades of collecting strains from around the world before the advent of modern sequencing, and the promiscuous way in which *R. toruloides* and related strains can mate producing viable progeny (Banno, 1967), has resulted in some confusion over the taxonomy of these, and related strains. Wang *et al.* (2015 a, b) recently analysed the phylogeny of the Basidiomycotina by sequence analysis revealing inconsistences in the taxonomy of *Rhodotorula*. Firstly, a number of strains, described as being a member of the genus *Rhodotorula* were found to be distantly related (*R. marina*, *R. lamellibrachiae*, and *R. armeniaca* renamed *Symmetrospora marina*, *Sakaguchia lamellibrachiae* and *Buckleyzyma armeniaca* respectively). As well as this reclassification of distant genera, strains in the closely related polyphyletic genera *Rhodosporidium*, *Rhodotorula*, *Sporobolomyces* and *Sporidiobolus* have been rearranged and renamed in accordance with the “One Fungus = One Name” nomenclature principle into three genera: *Rhodotorula, Rhodosporidiobolus* and *Sporobolomyces*. As *Rhodosporidium* and *Rhodotorula* were the teleomorph and anamorph names of the same genus, in line with modern fungal naming guidelines only the older name (*Rhodotorula*) was retained (Hawksworth, 2011). Thus *Rhodosporidium toruloides* was renamed *Rhodotorula toruloides* (Wang *et al.*, 2015b).

As well as this taxonomic confusion, strain variance plagues *R. toruloides*. Different groups around the world work with different strains, often exhibiting

significant genotypic differences. For example, comparison of *R. toruloides* strains CBS 14 and CBS 349 genomes found that while some regions are highly conserved, others are highly divergent, with a total genomic conservation of only 87 % (Kumar *et al.*, 2012, Paul *et al.*, 2014, Zhang *et al.*, 2016). This historic taxonomic confusion and strain variation can impede work as the relationship between strains described in literature is not necessarily clear and results obtained for one strain may not hold for another.

During this work I focused on the *R. toruloides* haploid strain CBS 14 as it is haploid, simplifying planed gene deletion experiments, its lipid production is well characterised (Evans & Ratledge, 1984a, Wiebe *et al.*, 2012, Zhang *et al.*, 2016), the genome has been sequenced (Kumar *et al.*, 2012, Zhang *et al.*, 2016), and it is almost identical to strain NP 11 (Zhu *et al.*, 2012, Zhang *et al.*, 2016) which has been the subject of in depth multi-omic study (Zhu *et al.*, 2012).

1.2.2 Physiology of lipid accumulation in *R. toruloides*

R. toruloides has a number of adaptions in its lipid biosynthesis machinery and core metabolism to facilitate fatty acid overproduction (Zhu *et al.*, 2012). Fatty acids are produced by the stepwise conjugation of two-carbon units from malonyl-CoA (produced by the carboxylation of acetyl-CoA) by fatty acid synthase (FAS). There are two classes of FAS: type I, found in animals, some fungi including *R. toruloides* and a minority of bacteria (which also possess a type II FAS); and type II, found in bacteria, plastids and mitochondria. Fatty acid biosynthesis requires the coordinated action of seven enzymatic activities: (1) activation of the FAS by conjugation of a phosphopantetheine

prosthetic group to the acyl carrier protein (ACP) domain; (2) priming by addition of the first acetate group to the to the attached phosphopantetheine; several rounds of extension where (3) an extending malonyl-CoA is loaded by the malonyl-palmitoyl transferase (MPT) domain (4) the fatty acyl chain is extended by condensation with malonyl-CoA, (5) the resulting β-ketoacyl-ACP is reduced to β-hydroxyacyl-ACP, (6) dehydrated to enoyl-ACP and (7) reduced to form an extended acyl-ACP; and finally, when a length of 16 or 18 carbons is reached the fatty acyl-CoA is released by the MPT domain (Jenni *et al.*, 2007). In type I FASs, all activities are performed by one protein consisting of one or two multifunctional polypeptides. Normally type I FASs have eight functional domains, including seven immobile catalytic domains and the mobile ACP domain that binds the growing fatty acid and passes it between the active sites, however *R. toruloides* FAS has two ACP domains (Zhu *et al.*, 2012, Fischer *et al.*, 2015). It is hypothesised that this increases the local concentration of intermediates around the active sites within the FAS, therefore increasing the rate of reaction. A second unique feature of the *R. toruloides* FAS is that this enzyme is made up of two polypeptides with a unique split between them, however this is not thought to influence the activity of the FAS (Fischer *et al.*, 2015).

As well as a unique FAS, in common with other oleaginous yeast *R. toruloides* has adaptations to facilitate overproduction of acetyl-CoA from which fatty acids are produced. Under lipid accumulating conditions *R. toruloides* accumulates high levels of the enzyme ATP citrate lyase (ACL) (Boulton & Ratledge, 1981a). This enzyme is found in the cytosol and in an ATP- and CoA-dependent manner cleaves citrate to oxaloacetate and acetyl-CoA (red

activity in Figure 1.3). This provides cytosolic acetyl-CoA which, via the action of the enzyme acetyl-CoA carboxylase provides malonyl-CoA as the substrate for fatty acid biosynthesis. The activity of ACL has been proposed as the rate-limiting step for lipogenesis in the oleaginous yeast *Lipomyces starkeyi* and this enzyme is absent in non-lipid-accumulating fungi (Boulton and Ratledge, 1981 a, b).

A second enzyme found in *R. toruloides*, and many other (but not all) oleaginous yeast is cytosolic $NADP^+$ reducing malic enzyme (Zhu *et al.*, 2012, Ratledge, 2014), responsible for the for the $NADP^+$-dependent decarboxylation of malate to pyruvate (purple in Figure 1.3). NADPH is required as the source of reducing potential for fatty acid biosynthesis, with two molecules of NADPH required per two-carbon unit added to the growing fatty acid. Under lipid accumulating conditions $NADP^+$ reducing malic enzyme, together with the enzymes pyruvate decarboxylase (which converts pyruvate to oxaloacetate) and malate dehydrogenase (responsible for the NADH-dependent reduction of oxaloacetate to malate), create a 'transhydrogenase cycle' responsible for transfer of reducing potential from NADH to $NADP^+$. Relative to other yeast such as *Y. lipolytica* where the pentose phosphate pathway is the sole source of cytosolic NADPH, the presence of $NADP^+$ reducing malic enzyme increases the theoretical maximum yield of lipid from glucose by 14 % (Ratledge, 2014).

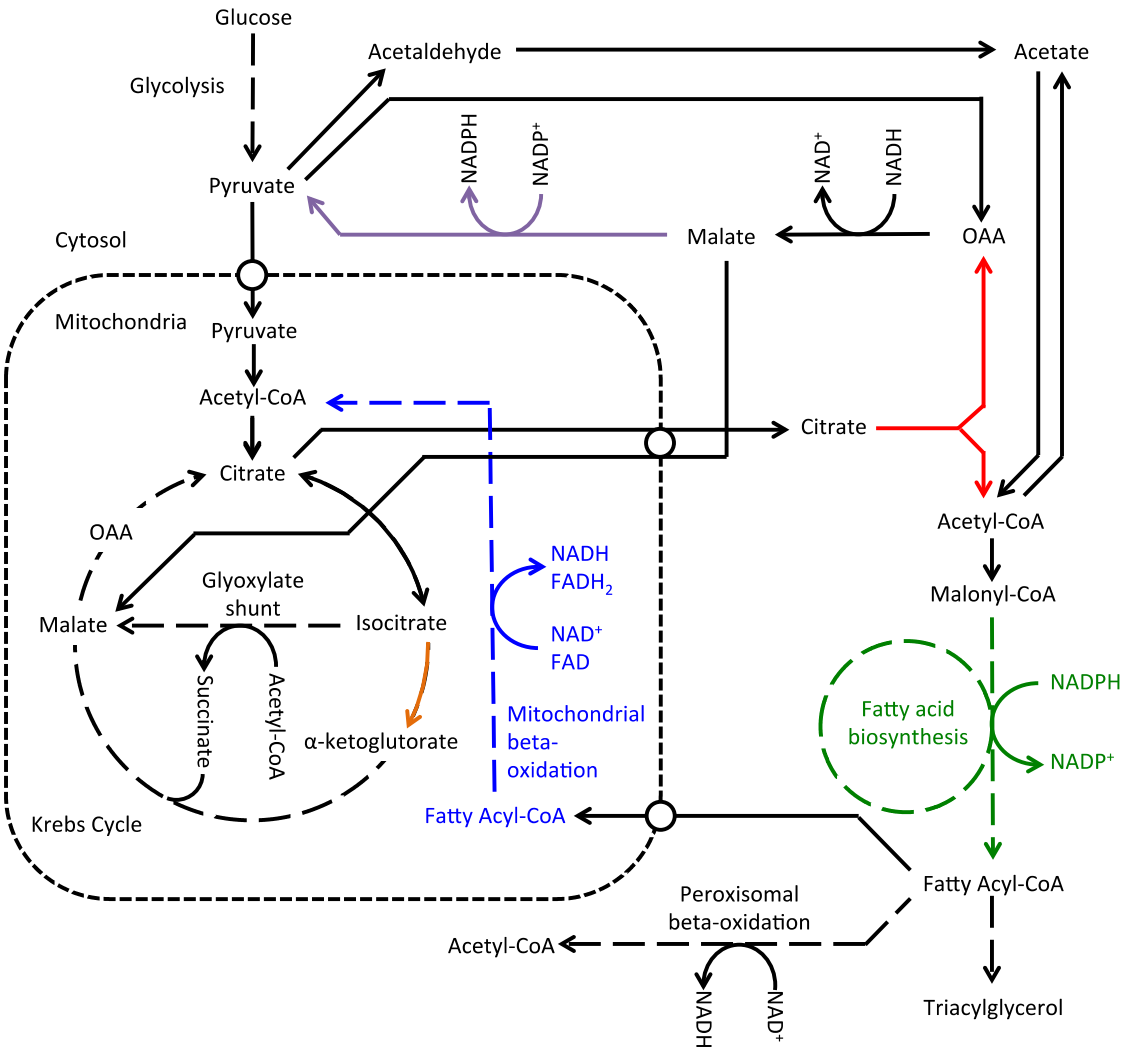


Figure 1.3. **Metabolic map highlighting key reactions for fatty acid biogenesis and degradation**. Solid arrows indicate individual reactions; pathways containing multiple reactions are indicated as dashed arrows. Adaptations of *R. toruloides* to life as an oleaginous yeast are highlighted, with fatty acid synthase in green, ATP citrate lyase in red, $NADP^+$ reducing malic enzyme in purple, isocitrate dehydrogenase in orange, and mitochondrial beta-oxidation in blue. Cofactors are indicated where significant for fatty acid overproduction. OAA = oxaloacetate.

As well as extra pathways to facilitate lipid overproduction, differences in regulation have a large impact on lipid accumulation. Citrate is produced in the Krebs cycle and is required for over production of acetyl-CoA. In the Krebs cycle, citrate undergoes a reversible aconitase-catalysed isomerisation to isocitrate, before decarboxylation by isocitrate dehydrogenase to 2-oxaloglutarate (orange in Figure 1.3). In oleaginous yeast, including *R. toruloides*, isocitrate dehydrogenase is allosterically regulated by AMP (Ratledge, 2004). Under low nitrogen conditions AMP is consumed by nitrogen scavenging AMP deaminase, thus the cellular AMP concentration is decreased, resulting in a decrease in the activity of isocitrate dehydrogenase. This results in a buildup of isocitrate, which reversibly equilibrates with citrate, thereby providing substrate for fatty acid biosynthesis.

Less relevant to its proposed use in industry, but nonetheless important to its oleaginous lifestyle, are adaptations of *R. toruloides* to use stored lipid. *R. toruloides* is able to metabolise lipid by both peroxisomal and mitochondrial beta-oxidation (blue in Figure 1.3) (Zhu *et al.*, 2012), the latter of which is lost in most fungi, including the Saccharomycotina (Poirier *et al.*, 2006, Gabriel *et al.*, 2014). With respect to the lipid, the reaction steps in perixosomal and mitochondrial beta-oxidation are the same, however during perixosomal beta oxidation, for every two-carbon unit metabolised, one acetate and one $NADH^{+}$ is released which can be used to produce 15 ATP; during mitochondrial beta oxidation, one acetate and one NADH is also released per two-carbon unit metabolised, but additionally it is also directly coupled to oxidative phosphorylation by the reduction of FAD to $FADH_2$; therefore, in mitochondrial beta oxidation up to 17 ATPs are produced per two-carbon unit metabolised.

1.2.3 Other potential uses of *R. toruloides*

As well as production of biodiesel, *R. toruloides* has been suggested as useful in other roles. In humans and other animals vitamin A_1 is essential for vision, regulation of the immune system and body patterning. Humans cannot produce vitamin A_1 *de novo*, however this can be produced *in vivo* by cleavage of many carotenoids creating a market for carotenoids as dietary supplements. To protect cells from light and oxidative damage *R. toruloides* accumulates carotenoids at concentrations of up to 1.3 g L^{-1}, principally torulene, γ-carotene, torularhodin, and β-carotene (Figure 3.2) (Buzzini *et al.*, 2007). These high yields make *R. toruloides*, or related yeasts, appealing for industrial carotenoid production (Frengova & Beshkova, 2009), however the most valuable carotenoid, β-carotene is produced in lesser amounts than the lower value torulene or other carotenoids. Therefore, for commercial production, one would ideally modify *R. toruloides* to maximise production of the more valuable carotenoids.

R. toruloides has been proposed as a bio-control agent. Application of *R. toruloides* to crops could reduce losses to disease and the amount of fungicide required to protect crops (Buck & Andrews, 1999), and application of *R. toruloides* to fruit post-harvest can result in a reduction in losses due to mould (ChandGoyal & Spotts, 1996, Filonow *et al.*, 1996). Saprotrophic yeasts such as *R. toruloides* achieve this protection primarily through nutrient and space competition with pest species, but also through induced resistance in the host (Lu *et al.*, 2014).

When grown with phenylalanine as a food source *R. toruloides* accumulates the enzyme phenylalanine ammonia lyase (PAL) to high concentrations (Gilbert & Tully, 1982). PAL catalyses the cleavage of phenylalanine to ammonia and trans-cinnamic acid facilitating the use of phenylalanine as both a carbon and nitrogen source. PAL is a proposed as a treatment for the hereditary human condition phenylketonuria (PKU) (Al Hafid & Christodoulou, 2015). In a healthy individual, excess dietary phenylalanine is converted to tyrosine by the action of the enzyme phenylalanine hydroxylase (PAH), after which it can be degraded, however in patents with PKU, PAH is lost which, if not managed can allow phenylalanine to accumulate to toxic concentrations causing neurological damage. The trans-cinnamic acid and ammonia produced by PAL can be metabolised *in vivo*, therefore PAL could potentially be used to detoxify accumulated phenylalanine. Unlike PAH, PAL is a monomeric enzyme and does not require cofactors for activity making it more suitable as a therapeutic (Al Hafid & Christodoulou, 2015). *R. toruloides* has previously been used for production of PAL, however while *R. toruloides* PAL has been demonstrated to reduce circulating levels of phenylalanine in animal disease models, the protein is immunogenic and rapidly degraded limiting its efficacy (Fritz *et al.*, 1976). Therefore were this enzyme be used as a therapeutic, it would be likely to be a decorated form, not produced in *R. toruloides* (Gamez et al., 2005).

1.3 Molecular genetics of *R. toruloides*

Relative to model yeasts or more traditional industrial yeasts, the molecular genetic tools required for manipulation of *R. toruloides* are lacking, however

some sporadic studies have been performed on this organism beyond those directly pertaining to its industrial application.

1.3.1 Life cycle

Sexuality and mating are a fundamental characteristic of an organism and a powerful tool for genetic manipulation. *R. toruloides* normally exists as vegetatively growing haploid yeast but is capable of a perfect fungal life cycle. *R. toruloides* has a bipolar mating system with mating types MAT A1 (A) (for example strain CBS 14) and MAT A2 (a) (for example strain CBS 349) (Banno, 1967). Haploid strains of opposing mating type can mate by conjugation giving rise to a dikaryotic mycelium. This dikaryotic mycelium can undergo karyogamy, giving rise to teliospores which, upon germination develop promycelium before undergoing meiosis giving rise to haploid yeast cells completing the life cycle. As well as this perfect life cycle, there are strains that deviate from this, for example the diploid strain CBS 6016 or the aneuploid strain CGMCC 2.1609, likely formed as a result of a failure of meiosis.

1.3.2 Prior molecular genetics

The earliest reported molecular genetic work on *R. toruloides* was related to production and characterisation of PAL. Gilbert *et al.* (1983) purified PAL mRNA, confirming its identity by expressing it in a rabbit reticulocyte lysate system, and used this to identify and clone the *R. toruloides* PAL gene (Gilbert *et al.*, 1985). Using this gene on a plasmid they reported stable and unstable transformation of *pal*$^-$ *R. toruloides* by protoplasting, performed by

degrading the yeast cell wall using lytic enzyme prepared from *Penicillium lilacinum*, and chemical transformation of spheroplasts. The *PAL* gene was used as a selectable marker for growth with phenylalanine as the sole carbon and nitrogen source. They also reported transformation of a *leu*$^-$ *R. toruloides* strain by selection with *S. cerevisiae LEU2* (Tully, 1985, Tully & Gilbert, 1985). No publication had subsequently reported successful repetition of this transformation protocol and effort has been put into repeating this work here in Exeter (Aves, personal communication) and elsewhere (Lin *et al.*, 2014) but this has been unsuccessful.

While the protoplasting transformation reported by Tully and Gilbert (1985) has not been subsequently reported, in 2012 Liu *et al.* demonstrated transformation of *R. toruloides* strain CBS 349 by *Agrobacterium tumefaciens*-mediated transformation (ATMT) and selection of transformats by growth on the antibiotic hygromycin B. This will be discussed further in chapter 3.

1.3.3 Genomics

Several genomes with different states of annotation have been published for *R. toruloides* and related species (Kumar *et al.*, 2012, Zhu *et al.*, 2012, Grigoriev *et al.*, 2014, Paul *et al.*, 2014, Zhang *et al.*, 2016), as well as corresponding transcriptomic and proteomic data. The *R. toruloides* genome is relatively compact, with a size of 20 megabases and around 8000 predicted genes. Compared with other members of the Pucciniomycotina this is small, for example the rusts have large genomes over 100 megabases in size, with over 25 000 genes (Zheng *et al.*, 2013, Tavares *et al.*, 2014), making an argument for *R. toruloides* as a model system for this fungal subphylum.

The *R. toruloides* genome has a GC content of 62 % and consequently a skewed codon usage with almost exclusive use of G or C at the wobble position. Therefore, for successful expression of transgenes in *R. toruloides* they must be codon optimised (Liu *et al.*, 2013); although the reason for this requirement has not been determined in *R. toruloides*, in *U. maydis* it is due to premature polyadenylation at AT-rich regions (Zarnack *et al.*, 2006).

In common with other basidiomycetes, the majority of *R. toruloides* genes contain multiple short introns, with transcripts containing a median of four introns per gene with a median length of 62 base pairs each, and a median exon length of 146 (Zhu *et al.*, 2012).

1.4 Project aims

The aim of this project was to develop molecular genetic tools for manipulation of *R. toruloides* CBS 14 and to validate this yeast as a system for production of hydrocarbons as a drop-in replacement petrodiesel. This first required development of protocols and selection markers to facilitate transformation of *R. toruloides* CBS 14. Using this transformation system it was hoped to develop a protocol for targeted gene deletion and a versatile, regulatable system for transgene expression. Finally, using the tools developed it was intended to express genes for hydrocarbon biosynthesis in *R. toruloides* as a proof of principle for using this yeast for hydrocarbon biosynthesis.

2 Materials and methods

2.1 Strains and media

2.1.1 *Rhodotorula toruloides* strains

R. toruloides wild-type haploid strains CBS 14 (*MAT-A1*; ATCC 10788, IFO 0559, MTCC 457; Rennerfelt, 1937), CBS 349 (*MAT-A2*; ATCC 10657, IFO 0880; Okunuki, 1931) and diploid type strain CBS 6016 (*MAT-A1*/*MAT-A2*; IFO 8766; Banno, 1967) were obtained from the Centraalbureau voor Schimmelcultures, Utrecht, The Netherlands. The CBS 349 derived haploid strain NCYC 1585 (*MAT-A2 leu2 ino*; Tully, 1985) was obtained from The National Collection of Yeast Cultures, Norwich, UK.

2.1.2 *R. toruloides* media and culture conditions

R. toruloides was grown at 30 °C in YPD (Sambrook and Russell, 2001), or Yeast Nitrogen Base without amino acids (YNB) (ForMedium, Hunstanton, UK) with 20 g L^{-1} glucose.

Promoter induction and repression media were YNB with 20 g L^{-1} glucose modified as follows: for *SGA1* induction medium, glucose was replaced with maltose; for *ICL1* and *ICL2* induction medium glucose was replaced with 200 mM sodium acetate; for *NAR1* induction medium, YNB without ammonium sulfate was supplemented with 0.78 g L^{-1} potassium nitrate; for *THI5* and *THI4* induction medium YNB without thiamine was used and for repression medium 20 mg L^{-1} thiamine was included; for *MET16*, 1 mM methionine was included in repression medium; for *CCC2* induction medium contained 20 µM $CuSO_4$

and repression medium was formulated without copper; for *CTR3* and *CTR31* initial screens induction medium was formulated without copper and repression medium contained 20 µM $CuSO_4$, for time course and promoter cut-down experiments induction medium contained 100 µM bathocuproinedisulfonic acid. Solid media contained 2 % agar except for *NAR1* induction/repression media where 2 % agarose was used.

For lipid accumulation, medium was YNB (ForMedium, Hunstanton, UK) modified to contain 30 g L^{-1} glucose and 1.74 g L^{-1} ammonium sulfate, giving a final C/N ratio of 65.

2.1.3 Other microbial strains and media

Agrobacterium tumefaciens strains GV3101 (van Larebeke *et al.*, 1974) and AGL-1 (Lazo *et al.*, 1991) were grown at 28 °C in LB (Cold Spring Harbor, 2006) containing rifampicin (50 µg mL^{-1}). Cloning was performed using *Escherichia coli* NEB5α (New England Biolabs, Ipswich, MA) grown in LB at 37 °C. In-yeast assembly was performed using *Saccharomyces cerevisiae* strain BY4742 (*MATα his3Δ leu2Δ lys2Δ ura3Δ*) (Brachman *et al.*, 1998) grown in YPD, or YNB with 20 g L^{-1} glucose and Complete Supplement Mix without uracil (ForMedium, Hunstanton, UK) for auxotrophic selection.

2.2 Chemicals

Chemicals were supplied by Sigma Aldrich (St Louis, MO) unless indicated. Synthetic DNA was produced by GeneArt (ThermoFisher, Waltham, MA).

2.3 General molecular biological techniques

2.3.1 Routine DNA manipulation

DNA manipulation was performed using standard techniques (Sambrook & Russell, 2001). Except where indicated, PCR was performed using Q5 DNA polymerase (New England Biolabs, Ipswich MA) with oligonucleotides purchased from Eurofins (Ebersberg, DE). Restriction digests were performed using high-fidelity restriction endonucleases (New England Biolabs, Ipswich, MA). Where used for cloning, PCR products were purified using a GeneJET PCR purification kit (ThermoFisher, Waltham, MA). Plasmid DNA was prepared from *E. coli* by alkaline lysis using a GeneJET plasmid miniprep Kit (ThermoFisher, Waltham, MA). DNA gel electrophoresis was performed using 8 g l^{-1} agarose gels prepared with 10 µg mL^{-1} ethidium bromide; electrophoresis was performed at 130 V for 45 min or until adequate separation was achieved and bands visualised by UV transillumination. Where necessary bands were extracted and purified using a QIAquick gel extraction kit (Quiagen, Venlo, NL). All cloning was verified by Sanger sequencing (Source Biosciences, Nottingham, UK). Details of plasmid construction are given in section 2.4.

2.3.2 *R. toruloides* genomic DNA extraction

R. toruloides genomic DNA was extracted using a yeast DNA extraction kit (ThermoFisher, Waltham, MA).

2.3.3 Bacterial transformation

E. coli was chemically transformed using high efficiency transformation (New England Biolabs, Ipswich, MA). Chemically competent *A. tumefaciens* was prepared and transformed by the protocol of Holsters et al. (1978).

2.3.4 Gibson assembly

Gibson assembly (Gibson *et al.*, 2009) was performed using a NEB Gibson Assembly Cloning Kit (New England Biolabs, Ipswich, MA).

2.3.5 Point mutation

Point mutation was performed using a Q5 site-directed mutagenesis kit (New England Biolabs, Ipswich, MA).

2.3.6 *S. cerevisiae* in-yeast assembly

In-yeast assembly was performed by a modified version of the protocol of Kilaru & Steinberg (2015). Briefly, DNA fragments with overlapping homology regions of 25 bp at their ends (1 µg total with fragments in stoichiometric amounts) were co-transformed into *S. cerevisiae* in an approximately equimolar ratio using a yeast transformation kit (Sigma-Aldrich, St Louis, MO). In-yeast assembled plasmids were extracted by the method of Singh and Weil (2002), transformed into *E. coli*, isolated by alkaline lysis and verified by Sanger sequencing of junctions.

2.3.7 Transformation of *R. toruloides*

Transformation of *R. toruloides* was performed using a modified version of the protocol of Liu *et al.* (2013). *A. tumefaciens* containing the appropriate binary plasmid was grown in LB with rifampicin (50 µg mL^{-1}) and kanamycin (50 µg mL^{-1}) at 28 °C for 48 h, then diluted to an OD of approximately 0.1 in induction medium (Gelvin, 2006) and incubated at 24 °C for 6 h. A 200 µL volume of this *A. tumefaciens* culture was then mixed with 200 µL of an overnight culture of *R. toruloides*, spread over a nitrocellulose membrane on solid induction medium and incubated at 24 °C for 48 h. Membranes were transferred to YPD with cefotaxime (150 µg mL^{-1}) and either G418 (150 µg mL^{-1}) or hygromycin (50 µg mL^{-1}) and incubated at 30 °C for 2-3 days. Colonies were restreaked to fresh selective YPD and grown overnight.

2.4 Plasmids constructed

2.4.1 *R. toruloides* transformation vectors

Plasmids pG418-Rt, pHyg-Rt and pGent-Rt were constructed using Gibson assembly. pCAMBIA0380 (Cambia, Canberra, AU) was digested using *Pvu*I. The *R. toruloides* CBS 14 *GPD1* promoter was PCR amplified from genomic DNA using primers RtGPD1F-pCambia0380 and either GPD1R-G418, GPD1R-Hyg, GPD1R-Gent (Table 2.1). The amplified promoter and digested backbone were assembled together with either a synthetic codon-optimised APH(3PH(3H(3.1). The am, hygromycin phosphotransferase or gentamicin-(3)-N-acetyltransferase gene (see appendix for sequences of synthetic DNAs) PCR amplified using primers pairs G418F-GPD1 and G418R-pCambia0380;

HygF-GPD1 and HygR-pCambia0380; or GentF-GPD1 and GentR-pCambia0380 respectively.

pEGFP-Rt-YR-G418 was constructed in two steps. First, plasmid pC-G418-YR (Sidhu *et al.*, 2015) was digested with *Pvu*II, and the codon-optimised G418 resistance gene and associated *GPD1* promoter (amplified from pG418-Rt using primers RtGPD1F-pCambia0380-2 and G418-NcTerm) was inserted by in-yeast assembly creating plasmid pG418-Rt-YR. This was digested using *Hin*dIII and assembled in-yeast with synthetic DNA comprising the *R. toruloides PGK1* promoter, codon-optimised EGFP gene, and CMV35S terminator.

pEGFP-Rt-YR-Hyg was constructed by digestion of plasmid pEGFP-Rt-YR-G418 with *Sma*I and insertion of the hygromycin resistance gene from plasmid pHyg-Rt amplified using primers HygF-GPD1-2 and Hyg-NcTerm, by in-yeast assembly.

pYFP-Rt-YR-G418 was constructed by digestion of plasmid pEGFP-Rt-YR-G418 with *Pml*I/*Spe*I and insertion of a synthetic Venus YFP by in-yeast assembly.

2.4.2 Plasmids for targeted integration

Plasmids pCrtI-Ko, pCrtY-Ko, pKu70-Ko and pKu80-Ko were produced for targeted deletion of genes *CRTI*, *CRTY*, *KU70* and *KU80* respectively. 1 kb regions flanking the target gene were PCR amplified from *R. toruloides* CBS 14 genomic DNA using primers listed in table 2.1, and inserted into *Eco*RI/*Hind*III-digested pG418-Rt by Gibson assembly.

Plasmid pKu80-5KbKO was constructed by PCR amplification of 5 kb fragments flanking the *KU80* gene using primer pairs Ku80-Up5kb-F/Ku80-Up-R and Ku80-Down-F/Ku80-Down5Kb-R, and insertion of fragments into *Eco*RI/*Hind*III-digested plasmid pG418-Rt-YR by in-yeast assembly.

For selection with carboxin, plasmid pCBX-Rt-YR-G418 was constructed; a synthetic *R. toruloides* CBS 14 *SDH2* gene with a c.671A>T point mutation was inserted into *Pml*I/*Spe*I-digested pEGFP-Rt-YR-G418 by in-yeast assembly. pCBX-Rt-YR was produced by in-yeast assembly of a 1 kb synthetic fragment comprising the 3' end of the *SDH2* gene including the c.671A>T point mutation and downstream intergenic DNA into *Eco*RI/*Afl*II-digested pEGFP-Rt-YR-G418 replacing the entire T-DNA region between the left and right border sequences.

For selection with leucine, plasmids pLeu2+ve and pLeu2-test were respectively constructed by PCR amplification of *R. toruloides* CBS 349 *LEU2* and upstream sequence using primers Leu2+Ve-F and Leu2+Ve-R, or 5' terminally truncated *R. toruloides* CBS 349 *LEU2* and downstream sequence using primers Leu2-Test-F and Leu2-Test-R, and insertion of each fragment into *Eco*RI/*Afl*II-digested pEFGP-Rt-YR-G418.

2.4.3 Plasmids for promoter analysis

Plasmids for testing promoter activity by EGFP expression were constructed by in-yeast assembly of *Afl*II/*Pml*I-digested pEGFP-Rt-YR-G418 with promoter fragments amplified from genomic DNA using respective primers (Table 2.1).

Plasmid pLeu-Rt-YR-G418 was constructed by amplification of the *R. toruloides* CBS 14 *LEU2* gene using primers LeuF-pCambia0380 and LeuR-CMV35S, and in-yeast assembly with pEGFP-Rt-YR-G418 digested with *Pml*I and *Spe*I. Plasmids for conditional *leu2* rescue under the regulation of *ICL1*, *NAR1*, *MET16* and *CTR3* 1500-bp promoter fragments were assembled in the same manner as plasmids for testing promoter activity with EGFP, with the modifications that pLeu-Rt-YR-G418 was used instead of pEGFP-Rt-YR-G418 as the base plasmid, and reverse primers Icl1R-Leu, Nar1R-Leu, Met16R-Leu and Ctr3R-Leu were used for amplification of *ICL1*, *NAR1*, *MET16* and *CTR3* promoters respectively.

2.4.4 Plasmids for hydrocarbon biosynthesis

For construction of plasmid pOleT-Rt-YR-G418 a codon optimised *Jeotgalicoccus sp.* ATCC 8456 *OleT* was synthesised and inserted by in-yeast assembly into *Pml*I/*Spe*I-digested pEGFP-Rt-YR-G418. A hexahistidine tag was subsequently added to the *OleT* gene by digestion of plasmid pOleT-Rt-YR-G418 with *Spe*I and in-yeast assembly with oligonucleotide OleT-H6-N to give plasmid pOleT-His6-Rt-YR-G418. Putative 5′ splice sites within the OleT gene were removed sequentially by site-directed mutagenesis using a Q5 site-directed mutagenesis kit (New England Biolabs, Ipswich, MA) with primer pairs Q5OleT1F and Q5OleT1R, and Q5OleT2F and Q5OleT2R.

Plasmid pLip2-Rt-YR-Hyg was produced by insertion of a synthetic, codon optimised *Thermomyces lanuginosus* lipase II gene inserted into *Pml*I/*Spe*I-digested pEGFP-Rt-YR-Hyg.

For alkane biosynthesis three synthetic gene constructs were produced: a codon optimised *Synechococcus elongatus* fatty acyl-ACP reductase; a codon optimised *Acinetobacter baylyi* fatty acyl-CoA reductase; and a synthetic gene cluster containing codon optimised *S. elongatus* aldehyde decarbonylase, *E. coli* ferredoxin and ferredoxin reductase under the regulation of the *R. toruloides* CBS 14 *TUB1*, *THI5* and *THI4* constitutive promoters respectively. The synthetic gene cluster had elements at both ends for in yeast assembly into *Afl*II/*Spe*I digested pEGFP-Rt-YR-G418 and were used to produce plasmid pDEC-G418. At the 5′ end this homology fragment was removed by digestion with *Eco*RI and a three part in-yeast assembly performed with the digested synthetic gene cluster, *Pml*I/*Spe*I-digested pEGFP-Rt-YR-Hyg and either the synthetic fatty acyl-ACP reductase gene or fatty acyl-CoA reductase gene, giving plasmids pDEC-AAR-G418 or pDEC-ACR-G418 respectively.

Table 2.1. Oligonucleotides used in plasmid construction

Primer name	Sequence[1]
RtGPD1F-pCambia0380	cacgtgtgaattacaggtgaccagctcgaatttccccgatCTGCAGAACTACGCCCTCTC
GPD1R-G418	tgcgtcttctccttgcccatTGTGAGTGATCTGGTGTTGTTC
G418F-GPD1	acaacaccagatcactcacaATGGGCAAGGAGAAGACGCA
G418R-pCambia0380	ttcaatcttaagaaactttattgccaaatgtttgaacgatcgCTAGAAGAACTCGTCGAGCATGAG
GPD1R-Hyg	gtgagctccggcttcttcatTGTGAGTGATCTGGTGTTGTTC
HygF-GPD1	acaacaccagatcactcacaATGAAGAAGCCGGAGCTCACCG
HygR-pCambia0380	ttcaatcttaagaaactttattgccaaatgtttgaacgatcgCTACTCCTTGGCGCGCGGGC
GPD1R-Gent	tcgttcgacgagcggagcatTGTGAGTGATCTGGTGTTGTTC
GentF-GPD1	acaacaccagatcactcacaATGCTCCGCTCGTCGAACGACG
GentR-pCambia0380	ttcaatcttaagaaactttattgccaaatgtttgaacgatcgCTACGTGGCCGTCGACGGGT
RtGPD1F-pCambia0380-2	ggcgcgccgaattcgagctcggtacccaaCTGCAGAACTACGCCCTCGC
G418-NcTerm	cagaggagcctgaatgttgagtggaatgatCTAGAAGAACTCGTCGAGCA
HygF-GPD1-2	CGCTCAGAACAACACCAGATCAGTCACA
Hyg-NcTerm	gagcctgaatgttgagtggaatgatCTACTCCTTGGCGCGCGGG
LeuF-pCambia0380	ctcacccgtccaactcccaccctcccacgtgcagcccaccATGCCCTACTCTATCACCTGCTTG
LeuR-CMV35S	ctactcacacattattatggagaaaactagtTCACTTCTTGGTAAGCAATCCCGT
ICL1-1500-F	gaccggcaacaggattcaatGTTCTACAAGGACGTTTGGC
ICL1-800-F	gaccggcaacaggattcaatGTCCTGCGCAGCGGCG
ICL1-600-F	gaccggcaacaggattcaatTGGTGCGTTCGCGTGCGT
ICL1-400-F	gaccggcaacaggattcaatGGACCGCATCCCGTGCGTC
ICL1-200-F	gaccggcaacaggattcaatACTTTGACTCGCATTACACTTTTTTCTCCGC
ICL1-100-F	gaccggcaacaggattcaatGGCTTTCTTTCTCTCTCTGCGAACGAGG
ICL1-R	ttcgagaccggatccgccatCTCGTGTGTAGTGTCGT
ICL2-1500-F	gaccggcaacaggattcaatCGCCGGCCGACCACCACTA
ICL2-R	ttcgagaccggatccgccatGGCGTGCACTCGTGACA
SGA1-1500-F	gaccggcaacaggattcaatCTCGGCAAGCACAGCTTGATG
SGA1R	ttcgagaccggatccgccatCGTGAGCGGGAGAGCG
NAR1-1500-F	gaccggcaacaggattcaatTGCGTCCGTCTCTCGGT
NAR1-800-F	gaccggcaacaggattcaatGTCTCCGCAGAATCGTCGGACC
NAR1-600-F	gaccggcaacaggattcaatAGCAGCTCTCGTCTTGTCGCTTGG
NAR1-400-F	gaccggcaacaggattcaatCAACGTCGGCCCGCCTTGT
NAR1-200-F	gaccggcaacaggattcaatCGGACAGCAACTCTGGCTCTGG
NAR1-100-F	gaccggcaacaggattcaatCGCTGGTCTTGTTGGACAGCTGG
NAR1-R	ttcgagaccggatccgccatTCTGCTAGTGCTGTAGGTG
THI5-1500-F	gaccggcaacaggattcaatTGCGTCCGTCTCTCGGT
THI5-R	ttcgagaccggatccgccatTCTGCTAGTGCTGTAGGTG
THI4-1500-F	gaccggcaacaggattcaatGCAGAGCAAGAAGAACC
THI4-R	ttcgagaccggatccgccatGTTGATTCTTAAACGTC
MET16-1500-F	gaccggcaacaggattcaatGCAAGGTGTTGGAGATGTC
MET16-800-F	gaccggcaacaggattcaatATAGAGCGCCATCTTCTCGAGC
MET16-600-F	gaccggcaacaggattcaatAGGCGGGCTGCTGAAGG

MET16-400-F	gaccggcaacaggattcaatCGGGCGTCGCAGGC
MET16-200-F	gaccggcaacaggattcaatCTGTGTGCGCCCGACTTG
MET16-100-F	gaccggcaacaggattcaatCGCGTGCTTCGCTCTTG
MET16-R	ttcgagaccggatccgccatCTGTTGAGGGTGCG
CCC2-1500-F	gaccggcaacaggattcaatCAGCGGAGTCTGTCGGTCGA
CCC2-R	ttcgagaccggatccgccatGGCGAACTCGGGCGA
CTR3-1500-F	gaccggcaacaggattcaatAGGTACTTGGAGAGGGCTGC
CTR3-800-F	gaccggcaacaggattcaatGGGCACGCGGAGGG
CTR3-600-F	gaccggcaacaggattcaatCGCAAAAACAGCGCATCC
CTR3-400-F	gaccggcaacaggattcaatTCTCCCAGCCGCTCCTCTAG
CTR3-200-F	gaccggcaacaggattcaatTGGGGTCGCTCTGAGGG
CTR3-100-F	gaccggcaacaggattcaatGCACGCAGCCTCAACCG
CTR3-R	ttcgagaccggatccgccatCGCGGATCGCAGAT
CTR31-1500-F	gaccggcaacaggattcaatGCGCAACGCACGGAGACC
CTR31-R	ttcgagaccggatccgccatCGTTCAGCAAGCGCACG
Icl1R-Leu	ccaagcaggtgatagagtagggcatCTCGTGTGTAGTGTCGT
Nar1R-Leu	ccaagcaggtgatagagtagggcatGTTCGTGGGTCGTTCTTC
Met16R-Leu	ccaagcaggtgatagagtagggcatCTGTTGAGGGTGCG
Ctr3R-Leu	ccaagcaggtgatagagtagggcatCGCGGATCGCAGAT
CrtI-Up-F	catgttgggcccggcgcgccgTCCACCTCTCAACCCACC
CrtI-Up-R	tgcagttgggtaccgagctcgCCTTGCTGTGCTAACGAG
CrtI-Down-F	gagtcgacctgcaggcatgcaTTGAGCGGGAGGAGGGAG
CrtI-Down-R	accatggtggactcctcttaaCCTCCTCCGCGCGTTCTT
CrtY-Up-F	catgttgggcccggcgcgccgGGATGAGGTGGAGAGACCAG
CrtY-Up-R	tgcagttgggtaccgagctcgGCGAGCGCGAGTCTAGCA
CrtY-Down-F	gagtcgacctgcaggcatgcaGTGACGGGGCAAAAGCTGGATCTTTAC
CrtY-Down-R	accatggtggactcctcttaaCGCTCCCGTTGCGCCGCT
Ku70-Up-F	catgttgggcccggcgcgccgCGTGGGTCGGCGAAGAAGG
Ku70-Up-R	tgcagttgggtaccgagctcgGCGATGAGGAGGACGACGACTG
Ku70-Down-F	gagtcgacctgcaggcatgcaGTATCTTCCAAACGATCGCGACATCCT
Ku70-Down-R	accatggtggactcctcttaaCCGCGATACTCGTTCGGCTTC
Ku80-Up-F	catgttgggcccggcgcgccgACTTCTAGCTCCGTCAAGGTTCGATG
Ku80-Up-R	tgcagttgggtaccgagctcgAGATGGAGGAGTGGGACCGCTTG
Ku80-Down-F	gagtcgacctgcaggcatgcaATTCCACCTAGTTCGTGCCTAGC
Ku80-Down-R	accatggtggactcctcttaaAGGTGCGTTCTGGCTTCC
Ku80-Up5kb-F	tggcaggatatattgtggtgtaaacaCCTCCTCGCCAACCTCGAAGAG
Ku80-Down5kb-R	aaacgctcttttctcttaggtttacTTACGATTCGCTCGTCGCAC
Leu2+Ve-F	ggcaggatatattgtggtgtaaacaGGAGCCGTTCGTATCGAGTGAGTT
Leu2+Ve-R	aaacgctcttttctcttaggtttacTCTTTGCCTGTGCTCGCAAAG
Leu2-Test-F	ggcaggatatattgtggtgtaaacaAAGCGCGCCCGTGCTGT
Leu2-Test-R	aaacgctcttttctcttaggtttacCAGGTCCTCCCGCGACGA
OleT-H6-N	CTCACCCGTCCAACTCCCACCCTCCCACGTGCAGCCCACCATGCACCACC ACCACCACCACGCCACCCTCAAGCGCGACAAGGGCCT
Q5OleT1F	CCATTTGTGGTCGTCACCGGCAAGGAGGG
Q5OleT1R	CTTGCCGCCGAGGGCCTT
Q5OleT2F	CCATTTGTTCCTTTCCTCCCGGGCAAGGCCA
Q5OleT2R	GTAGTAGCGGCGGACCTCCTGG

[1] Priming sequences are shown in uppercase and 5' extensions in lowercase.

2.5 Whole cell analytical methods

2.5.1 Auxotrophic rescue

Cells were grown overnight in induction or repression medium with leucine (100 mg L^{-1}) as indicated, harvested by centrifugation, washed twice with sterile water and re-suspended in sterile water to approximately 10^6 cells mL^{-1}. 10 × serial dilutions were then spotted on to solid induction and repression media with or without leucine using a replica plater (Sigma-Aldrich, St Louis, MO).

2.5.2 Determination of antibiotic minimum inhibitory concentration by agar dilution assay

Agar dilution was performed by the protocol of Wiegland *et al*. (2008). Cultures were grown overnight in non-selective YPD and 10 × serial dilutions of cells were spot plated to YPD agar with 20, 10, 5, 2, 1, 0.5 or 0 µg ml^{-1} antibiotic and incubated at 30 °C for 24 h.

2.5.3 Microscopy for measurement of EGFP expression

R. toruloides expressing EGFP was grown overnight in YNB and observed under bright field and fluorescence conditions using an Olympus IX81 optical microscope (Olympus, Tokyo, JP) equipped with a 470/40 nm ET bandpass excitation filter, a T 495 nm LPXR beamsplitter and a 525/50 nm ET bandpass emission filter (Chroma Technology, Olching, DE). Fluorescence intensity was quantified using ImageJ (Abramoff *et al.*, 2004).

2.5.4 Flow cytometry for measurement of EGFP expression

For initial screening and promoter cut-down experiments three independent transformants were each grown overnight in YNB, pelleted by centrifugation (2500 *g* for 5 min) and washed twice with sterile water. Approximately 10^7 cells were added to 20 mL induction/repression medium and allowed to grow for 16 h (8 h for *MET16* promoter cut-down experiments). Samples of 0.5 ml were then taken and kept on ice until fluorescence could be measured. For measurement of induction rates, starter cultures were grown overnight in repressive conditions. Cells were then pelleted by centrifugation and washed twice with sterile water. Approximately 10^7 cells were added to 50 mL induction or repression medium and grown at 30 °C. Samples of 0.5 mL were taken at the indicated time intervals and kept on ice until fluorescence could be measured.

Fluorescence was quantified by flow cytometry using a FACSAria II (BD Biosciences, San Jose CA) with excitation at 488 nm and a 530/30 nm emission filter. To quantify cell density, CountBright absolute counting beads (ThermoFisher, Waltham, MA) were added to samples. Data were analyzed using FlowJo software (FlowJo, Ashland, OR) to determine median fluorescence for each sample. Student's t-tests were conducted to determine statistical significance between different experimental conditions.

2.6 Molecular biological analytical methods

2.6.1 Thermal asymmetric interlaced PCR (TAIL-PCR)

TAIL-PCR was performed according to Zhou *et al.* (2009). Transformant cultures were grown overnight and genomic DNA extracted. PCR amplifications were performed using thermal conditions and templates listed in table 2.2 with primers indicated in table 2.3. In each case the reaction mixture used was 2 × GO-TAQ DNA polymerase master mix (Promega, Fitchburg, WI) (10 µl), ultrapure water (8.8 µL), 100 µM arbitrary degenerate primer (0.5 µL), 10 µM specific primer (0.2 µL) and template (0.5 µL). After the final round of PCR, products were separated by gel electrophoresis and the brightest band excised. The DNA was purified and Sanger sequencing performed.

Table 2.2. Reaction conditions for TAIL-PCR

Round	Template	Specific primer	Conditions	
1	Genomic DNA	Tail-L1	1 ×	94 °C, 5 min
			5 ×	94 °C, 30 s 68 °C, 30 s 72 °C, 2 min
			1 ×	94 °C, 30 s 30 °C, 3 min ramping to 72 °C at 0.3 °C s^{-1} 72 °C, 2 min
			15 ×	94 °C, 30 s 68 °C, 30 s 72 °C, 2 min 94 °C, 30 s 68 °C, 30 s 72 °C, 2 min 94 °C, 30 s 50 °C, 1 min 72 °C, 2 min
			1 ×	72 °C, 7 min
2	50 × diluted primary reaction product	Tail-L2	1 ×	94 °C, 5 min
			15 ×	94 °C, 30 s 68 °C, 30 s 72 °C, 2 min 94 °C, 30 s 68 °C, 30 s 72 °C, 2 min 94 °C, 30 s 50 °C, 1 min 72 °C, 2 min
			1 ×	72 °C, 7 min
3	50 × diluted secondary reaction product	Tail-L3	As per round 2	

Table 2.3. Primers used in TAIL-PCR assembly

Primer	Sequence[1]
Tail-L1	CGTGGTGGTGGTGGTGGTGGCTAG
Tail-L2	ATGGTGGACTCCTCTTAAAGCTTGGCTGC
Tail-L3	GCGTTAATTCAGTACATTAAAAACGTCCGCAATG
GACD1	NYCGASCKTSGWGCT
GACD2	GTSGRCWGRSMCGSAT
GACD3	TGYGSAGYASCRSMGA
GACD4	TGCGNSGWMSCRSAG
GACD5	AGWGISGSMNCSWGG
GACD6	CAWCGSCNGWSRSGT
GACD7	TCSGICGNACISKSGA
GACD8	GTTSIKCSWGCWNSGC
GACD9	TCRGSYGWCIGSNSTG
GACD10	TCTYICGSRCSWNGGA
GACD11	TGSWGNGCIRSWCG
GACD12	GASYGWCSRGWGNSTC

[1] Standard IUPAC and IUB codes are used.

2.6.2 Western blotting

To determine whether hexahistidine-tagged OleT was expressed, western blotting was performed. 5 mL cultures were grown overnight and cells pelleted by centrifugation (2500 *g*, 5 min). The pellet was re-suspended in 1 mL lysis buffer (8 M urea, 65 mM DTT, 0.1% Triton X-100, 100 mM NaCl, 50 mM Tris–Cl, 1 mM PMSF, 1 mM EGTA, 1 mM EDTA, pH 7.4; (Liu *et al.*, 2009)) and cells lysed by bead beating using a Fast prep 24 (MP biomedicals, Santa Ana, CA) with lysing matrix C (1 mm glass beads; MP biomedicals, Santa Ana, CA); lysing conditions were 6 m s^{-2} for 4 × 30 s with samples maintained on ice for 5 min between treatments. SDS PAGE was performed using NuPAGE Novex 4-12 % Bis-Tris Protein Gel (ThermoFisher, Waltham, MA). Blotting was performed to a polyvinylidene fluoride membrane using a Pierce power blotter (ThermoFisher, Waltham, MA) and antibody treatment used an iBind western system (ThermoFisher, Waltham, MA). The primary antibody used

was rabbit anti-hexahistidine (ThermoFisher, Waltham, MA) and the secondary antibody was horseradish peroxidase-conjugated goat anti-rabbit (ThermoFisher, Waltham, MA). Blots were visualised by treatment with 3,3′,5,5′-tetramethylbenzidine.

2.7 Analytical methods for lipid and hydrocarbon analysis

2.7.1 Growth of *R. toruloides* for lipid accumulation and hydrocarbon biosynthesis

Starter cultures were grown overnight in 5 ml YPD containing, where appropriate, 50 µg mL^{-1} G418 or hygromycin. Cells were pelleted by centrifugation (2500 *g* for 5 min) and washed twice with sterile water. Approximately 10^7 cells were added to 50 mL lipid accumulation medium and grown at 30 °C with shaking for 72 hours.

2.7.2 Hydrocarbon extraction and measurement by GC-MS

Hydrocarbon extraction was performed using a modified version of the protocol of Chen et al. (2015). After growth in lipid accumulating conditions, cells were harvested by centrifugation, washed twice with ultrapure water and resuspended in 1 mL methanol with 0.1 % tetrabutylammonium hydroxide (TBAH) and either 1-tetradecene (25 ng $µL^{-1}$) when screening for alkenes or tetradecane (25 ng $µL^{-1}$) when screening for alkanes. Cells were lysed by bead beating using a Fast prep 24 (MP biomedicals, Santa Ana, CA) with lysing matrix C (1 mm glass beads; MP biomedicals, Santa Ana, CA); lysing conditions were 6 m s^{-2} for 10 × 30 s with samples maintained on ice for 5 min between treatments. Samples were transferred to glass vials with 1 mL

hexane and were mixed overnight at 4 °C. Samples were centrifuged (2500 *g* for 20 min) and the organic (upper) phase collected for GC-MS analysis. GC-MS was performed using an Agilent 7200B GC / quadrupole – time of flight (Q-TOF) mass spectrometry system (Agilent technologies, Santa Clara, CA). 0.8 µl sample was used for GC-MS analysis with a 10:1 split ratio. GC analysis was performed using a Phenomen 7HG-G027-11-GGC capillary GC column (30 m x 0.25 mm x 0.25 µm coating thickness; Phenomenex, Torrance, CA). Initial oven temp was 70 °C for 4 min, before increasing linerally to 310 °C at a rate of 15 °C s^{-1} and held for 6 min. The carrier gas used was He at a flow rate of 1.5 ml min^{-1}. MS was performed with an emission current of 35 µA, an emission voltage of 70 eV, an acquisition rate of 5.0 spectra s^{-1} and a spectrum acquisition time of 200 ms. Tetradecane, 1-tetradecene, pentadecane, 1-pentadecane, heptadecane and 1-heptadecene standards were analysed for peak identification and quantification.

2.7.3 Free fatty acid extraction and measurement by GC-FID

Measurement of free fatty acids was performed as per the protocol of Brown (2016). After growth in lipid accumulating conditions 1 mL of culture was taken and lysed by bead beating using a Fast prep 24 (MP biomedicals, Santa Ana, CA) with lysing matrix C (1 mm glass beads; MP biomedicals, Santa Ana, CA); lysing conditions were 6 m s^{-2} for 10 × 30 s with samples maintained on ice for 5 min between treatments. 100 µL sample was taken and 1 µL of a 5 g L^{-1} heptadecanoic acid internal standard added. 10 µL 40 % TBAH was added and samples incubated at 40 °C with shaking for 5 min. Ethyl acetate (50 µL) and iodomethane (50 µL) were added and incubated at 40 °C with

shaking for 30 s. Samples were centrifuged (2000 *g*, 1 min) and 50 µL of the organic (lower) phase collected. The solvent was removed by evaporation *in vacuo* and the dried samples re-suspended in 50 µL dichloromethane. 2 µL of each sample was used for GC-FID analysis with a 10:1 split ratio. GC-FID was performed using an Agilent 7890B system (Agilent technologies, Santa Clara, CA) equipped with a Phenomen 7HG-G027-11-GGC capillary column (30 m x 0.25 mm x 0.25 µm coating thickness; Phenomenex, Torrance, CA) and a flame ionising detector (Agilent technologies, Santa Clara, CA). Initial oven temp was 40 °C for 2.45 min, before increasing linearly to 310 °C at a rate of 24.522 °C s^{-1} and held for 4.08 min. The carrier gas used was H_2 with a flow rate of 1.2 ml min^{-1}. Palmitic (C16:0), palmitoleic (C16:1), stearic (C18:0), oleic (C18:1) linoleic (C18:2) and tetracosanoic acid (C24:0) acid standards were analysed for peak identification and quantification.

2.7.4 Dry cell weight

For measurement of dry cell weights 10 mL culture was placed in a pre-weighed tube. Cells were pelleted by centrifugation (2500 *g*, 5 min) and the culture medium removed. The pellets were dried in an oven at 100 °C for 24 hours and weighed. Dry cell weight was calculated as the final mass of the tube and dry pellet minus the mass of the tube.

2.7.5 Gravimetric measurement of total lipid content

Total lipid content was determined gravimetrically by a modified version of the protocol of Wiebe *et al*. (2012). After growth in lipid accumulating conditions, 10 mL culture was taken and cells pelleted by centrifugation (2500 *g*, 5 min),

washed twice with ultrapure water and resuspended in 1 mL methanol with 0.1 % TBAH. Cells were lysed by bead beating using a Fast prep 24 (MP biomedicals, Santa Ana, CA) with lysing matrix C (1 mm glass beads; MP biomedicals, Santa Ana, CA); lysing conditions were 6 m s^{-2} for 10 × 30 s with samples maintained on ice for 5 min between treatments. Cold chloroform (1 mL) and acetic acid (300 µl) were added and samples mixed vigorously for 10 min. Samples were centrifuged (2500 *g*, 20 min) and the organic (lower) phase was collected and placed in a pre-weighed tube. The aqueous (upper) phase again washed with 1 ml cold chloroform and combined organic fractions were pooled and the solvent removed *in vacuo*. Samples were weighed and lipid content calculated as the final mass of the tube and lipid minus the mass of the tube.

2.8 Bioinformatics

2.8.1 Motif discovery

De novo motif discovery was performed using MEME (Bailey *et al.*, 2009). For interspecies motif discovery, MEME version 4.11.2 hosted at http://meme-suite.org/tools/meme was used with default settings. Genomes used for comparison were: *R. toruloides* CBS 14 (Kumar *et al.*, 2012), *R. toruloides* CBS 349 (Zhang *et al.*, 2016), *R. graminis* WP1 (Firrincieli *et al.*, 2015), *Sporobolomyces* (formerly *Sporidiobolus*) *salmonicolor* CBS 6832 (Coelho *et al.*, 2015), *Sporobolomyces roseus* JGIBAIF-5F1, *Phyllozyma* (formerly *Sporobolomyces*) *linderae* CBS 7893, *Microbotryum lychnidis-dioicae* p1A1 (Grigoriev *et al.*, 2014), *Mixia osmundae* IAM 14324 (Toome *et al.*, 2014), *Leucosporidium creatinivorum* (formerly *Leucosporidiella creatinivora*)

(Grigoriev *et al.*, 2014) and *Puccinia graminis* (Duplessis *et al.*, 2011). For searching all *R. toruloides* promoters MEME version 4.11.2 hosted on a private server was used with default settings with the exception that 'maxsize' increased to 5 000 000. Searching for known elements within promoters was performed using FIMO, version 4.11.2 (Grant *et al.*, 2011) hosted at http://meme-suite.org/tools/fimo using standard settings.

2.9 Statistical analysis

2.9.1 Flow cytometry data

For analysis of the effect of introns on EGFP fluroescence, three independent transformant clones with each EGFP were grown and fluorescence was measured by flow cytometry and the median cellular fluorescence determined for ≥500 000 cells per culture. The mean of, and the standard deviation between the three median fluorescence values for each construct were then calculated and reported. A one way ANOVA assay was performed to determine if the median fluorescence values of any EGFP construct were significantly changed.

For analysis of inducible promoter activity, in each case three independent transformant clones were grown and each tested under both induced and repressed conditions. Cellular fluorescence was measured by flow cytometry and the median cellular fluorescence determined for ≥500 000 cells per culture. The mean of, and the standard deviation between the three median fluorescence values for each condition (induced or repressed) were then calculated and reported. A paired Students T-test was then performed to

determine the significance of any measured difference between the median fluorescence values under induced and repressed conditions.

2.9.2 Lipidomic data

For comparison of dry cell weight, total lipid content and free fatty acids after expression of lipase, three independent transformant clones, and three different wild type colonies were picked, grown and measured. The mean of, and standard deviation between the three replicates of each set was reported and an unpaired Students T-test performed to determine the statistical significance of any difference observed between the two.

3 DNA transformation and genomic integration of *R. toruloides* CBS 14

3.1 Introduction

Transformation and transgene expression are cornerstones of genetic engineering, and therefore are the first challenges for manipulation of *R. toruloides* CBS 14. Transformation can be considered as three problems: first, exogenous DNA must be introduced into viable cells, second, transformed cells need to be selected, and finally a way of replicating and maintaining the exogenous DNA is required. Many solutions to these challenges are available for commonly-used model organisms, for example *S. cerevisiae* can be transformed by treatment with lithium acetate and polyethylene glycol (PEG), selecting for transformants by either dominant or auxotrophic selection, with exogenous DNA maintained either as extrachromosomal plasmids or through chromosomal integration (Sambrook & Russell, 2001). However, the cell walls of other fungi are often much tougher than those of *S. cerevisiae*, and lithium acetate/PEG mediated transformation is ineffectual for transformation of *R. toruloides* (Abbott *et al.*, 2013). Also, molecular genetic elements for selection of transformants and maintenance of exogenous DNA in one species may not work in another; even when comparing fungi within the Pucciniomycotina, protocols developed for transformation of one species do not necessarily work in related organisms, with the efficacy of tools not necessarily following taxonomic relationships (Abbott *et al.*, 2013).

Prior to starting this work, two groups had reported transformation of *R. toruloides*: Tully and Gilbert (1985) (described in chapter one) and Liu *et al.* (2013). The methods for introduction of exogenous DNA, selection of transformants and maintenance of introduced DNA were different in each case. Tully and Gilbert (1985) described protoplast transformation of *R. toruloides*, followed by auxotrophic selection, whereas Liu *et al.* (2013) described *Agrobacterium tumefaciens*-mediated transformation (ATMT) of *R. toruloides* strain CBS 349 and selection of transformants by growth in the presence of the antibiotic hygromycin B (hereafter referred to as hygromycin). While the protocol of Tully and Gilbert is ostensibly simpler, no other groups have reported successful repetition of this work, and attempts to repeat this work here in Exeter and elsewhere have failed (Aves, personal communication; Lin *et al*., 2014). As a result it was decided to focus on ATMT.

3.1.1 *Agrobacterium tumefaciens*-mediated transformation

A. tumefaciens is a Gram positive bacterium and is the causal agent in plant crown gall disease (Smith & Townsend, 1907). During pathogenesis, *A. tumefaciens* inserts DNA into plant cells mediating production of opines which the bacteria use as a food source, and causing the formation of tumours in which the bacteria reside (Chilton *et al.*, 1977). Bacterial virulence is mediated by Ti-plasmids (van Larebeke *et al.*, 1974) which have two parts, the T-DNA region containing the genes transferred to plant cells and the backbone containing other bacterial virulence genes (Chilton *et al.*, 1978). The T-DNA region is delimited by two short, imperfect repeats known as the left and right borders (Yadav *et al.*, 1982, Zambryski *et al.*, 1982). It was realised that the

genetic material in the T-DNA region could be freely manipulated and could be used to shuttle any gene of interest into plants (Bevan *et al.*, 1983, Caplan *et al.*, 1983, Fraley *et al.*, 1983). Furthermore, the organisation of genes in the plasmid is unimportant and the T-DNA can be hosted on a shuttle vector, allowing simple manipulation of T-DNA in a tractable system such as *Escherichia coli*, and maintenance of virulence genes on a second helper plasmid, simplifying cloning (Hoekema *et al.*, 1983).

ATMT is summarised in Figure 3.1. Briefly, a binary plasmid, with material to be transferred flanked by the left and right border sequences, is transformed into *A. tumefaciens*. The transformed *A. tumefaciens* is induced to express virulence genes by conditions mimicking a plant wound: low pH (~5.6) (Mantis & Winans, 1992), inducing sugars such as glucose (Cangelosi *et al.*, 1990, Shimoda *et al.*, 1990), reduced temperature (~24 °C) (Fullner & Nester, 1996), and plant phenolic compounds induced by wounding (most commonly acetosyringone) (Stachel *et al.*, 1986). This leads to excision of single stranded T-DNA from between the left and right border sequences (Tinland *et al.*, 1994, Yusibov *et al.*, 1994) which associates with *A. tumefaciens* virulence proteins, including the VirD2 protein to which it covalently binds (Herrera-Estrella *et al.*, 1988, Vogel & Das, 1992). The *A. tumefaciens* cell associates with the cell to be transformed and injects the single stranded T-DNA protein complex by means of a type IV secretion system (Christie, 1997, Vergunst *et al.*, 2000). Once in the recipient cell, the T-DNA is targeted to the nucleus by a nuclear localisation signal in the VirD2 protein (Herrera-Estrella *et al.*, 1990, Howard *et al.*, 1992, Koukolikova-Nicola & Hohn, 1993), and the DNA is inserted into the chromosome. The mechanism by which the T-DNA is

inserted into the host chromosome is unclear and involves interaction of *Agrobacterium* and host factors, but the result is usually single insertion of T-DNA at an indeterminate locus in the host chromosome (Gelvin, 2003, Gelvin, 2010, van Kregten *et al.*, 2016).

Initially ATMT was used for transformation for plants; however in 1995 Bundock *et al.* demonstrated transformation of *S. cerevisiae*, and later, de Groot *et al.* (1998) demonstrated ATMT of the filamentous fungus *Aspergillus awamori* and six other fungal species including the basidiomycete *Agaricus bisporus*. It is difficult to directly compare transformation rates achieved by ATMT relative to other transformation methodologies as the commonly used metric of colony forming units per µg DNA is immaterial as no free DNA is used during ATMT and transformation rates will vary from organism to organism. However, when comparing transformants per number of recipient cells ATMT of *A. awamori* was found to be up to 600-fold more efficient than chemically mediated transformation (de Groot *et al.*, 1998). De Groot *et al.* (1998) observed that whilst transformation was successful with all strains, transformation rates varied greatly, with no apparent pattern between phylogeny and transformation rate. For example transformation of *A. awamori* was observed at a rate approximately 1000 × greater than that of *A. nidulans*. Since then ATMT has been demonstrated of many other fungal species including *R. toruloides* CBS 349 and other species recalcitrant to transformation by other methodologies (Michielse *et al.*, 2005, Abbott *et al.*, 2013, Liu *et al.*, 2013), and even human cells (Kunik *et al.*, 2001). ATMT of different fungal species or tissues generally uses similar protocols and the outcome is generally the same with single integration of T-DNA at an

indeterminate locus in the host chromosome (de Groot *et al.*, 1998, Michielse *et al.*, 2005, Abbott *et al.*, 2013, Liu *et al.*, 2013). However occasionally modifications to the protocol are required, for example, co-incubation of *A. tumefiaciens* and fungi is normally performed with yeast like cells or condia on solid media, but transformation of fruiting bodies of *A. bisporus* required vacuum infiltration of tissues (Chen *et al*., 2000). Also, a minority of fungi appear to be recalcitrant for ATMT e.g. *Sclerotinia sclerotiorum* (Rolland et al., 2003).

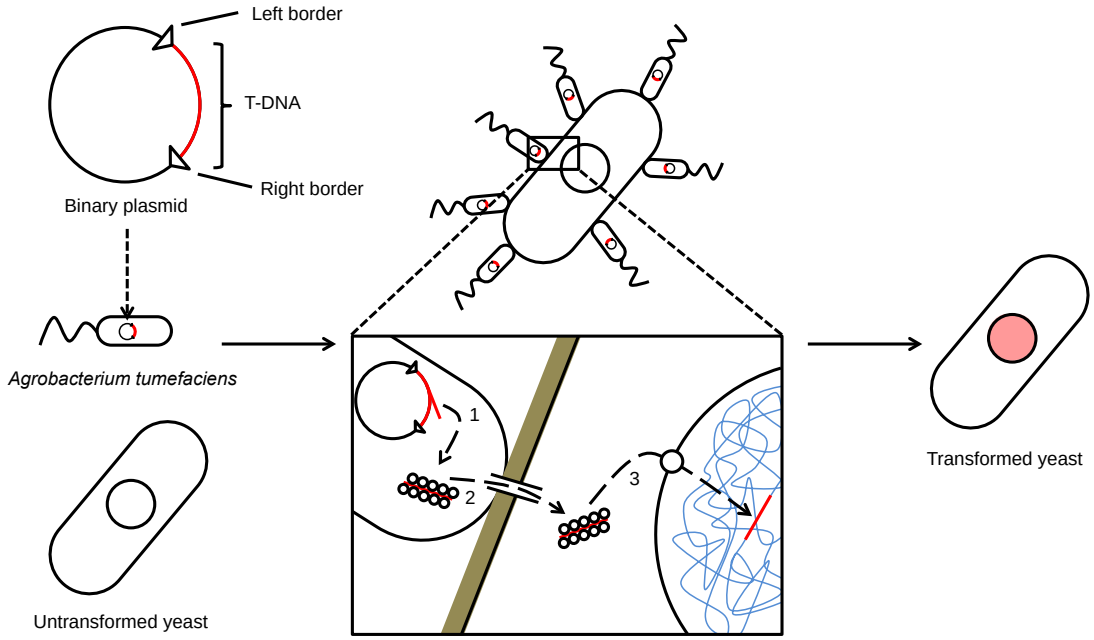


Figure 3.1. *A. tumefaciens*-mediated transformation of yeast. *A. tumefaciens* harbouring a binary plasmid including T-DNA (red) is mixed under inducing conditions with the yeast to be transformed. The *A. tumefaciens* attaches to the yeast cells and transfers T-DNA to the yeast chromosome. The exact mechanism of transformation is unclear, however (1) a single stranded copy of the T-DNA is excised and is bound by *Agrobacterium* virulence proteins including the VirD2 protein. (2) This DNA/protein complex is transferred to the recipient cell by a type IV secretion system. (3) The T-DNA is shuttled to the recipient nucleus through a nuclear pore complex due to the nuclear localisation signal on the VirD2 protein, and is inserted into the host chromosome. Antibiotic is applied to kill remaining *A. tumefaciens* cells and selection is imposed to kill untransformed yeast cells leaving only the transformed yeast.

3.1.2 Selection of transformants

While ATMT provides a solution to the issues of introduction and maintenance of DNA in *R. toruloides* CBS 14, a method is required to select for transformants. There are two alternative methods of selection, auxotrophic selection and dominant selection. The dominant antibiotic selection used by Liu *et al.* (2013) for transformation of *R. toruloides* CBS 349 has the advantage that it does not require prior generation of auxotrophs as a background for selection; however this does require expression of a heterologous antibiotic resistance gene. Liu *et al.* (2013) initially used *Streptomyces hygroscopicus* hygromycin phosphotransferase (*HPT*), under the regulation of the *R. toruloides* glyceraldehyde-3-phosphate dehydrogenase (*GPD1*) promoter; promoters from orthologues of this gene are used for constitutive expression of transgenes in other fungal systems (Punt *et al.*, 1990, Kuo *et al.*, 2004, Neveu *et al.*, 2007). Initial attempts at transformation and selection with hygromycin failed. They hypothesised that because of the high genomic GC-content, and the resulting codon bias of *R. toruloides* genes, codon optimisation would be required for successful transgene expression. When transformation was repeated with a codon optimised *HPT* gene, hygromycin resistant transformant *R. toruloides* colonies were observed (Liu *et al.*, 2013).

As generation of auxotrophic strains would likely be time consuming and there are no publicly available *R. toruloides* CBS 14 auxotrophs, and antibiotic selection has been demonstrated in *R. toruloides* CBS 349, it was decided to use antibiotic selection for transformation of *R. toruloides* CBS 14.

3.1.3 Targeted integration of T-DNA

In plants, animal cells and most fungi, ATMT results in apparently random integration of T-DNA into the chromosome by means of illegitimate recombination (Gelvin, 2003, van Attikum & Hooykaas, 2003, Gelvin, 2010, Kleinboelting *et al.*, 2015). Ideally T-DNA would insert at a predetermined locus as this would facilitate targeted gene deletion/disruption, would prevent positional effects causing differential expression of transgenes between transformant lines and prevent off-target effects resulting from insertional mutagenesis by integration of T-DNA into other genes (Matzke & Matzke, 1998).

Although integration of T-DNA into the chromosome during ATMT is generally random, in some cases T-DNA can be targeted to specific loci by homologous integration (Bundock *et al.*, 1995). This is because, during ATMT, DNA double strand break repair pathways are recruited and are responsible for much of the process of T-DNA integration (Tzfira *et al.*, 2003, van Attikum & Hooykaas, 2003). Therefore in cases where homologous recombination is the favoured double strand break repair pathway this is also the favoured pathway for integration of T-DNA, for example as is the case in *S. cerevisiae*. It was hoped that it would be possible to use homologous integration to target T-DNA to specific loci within the *R. toruloides* genome and four genes were initially targeted for deletion by homologous recombination: *CrtI*, *CrtY*, *KU70* and *KU80*.

3.1.4 Carotenoid production in *R. toruloides*

When grown to late log or stationary phase *R. toruloides* accumulates carotenoids, most significantly the red pigment torulene, which is responsible for the red colour of yeast colonies. Carotenoids are produced from geranylgeranyl pyrophosphate by the action of two enzymes, lycopene beta-cyclase (*CrtY*) and phytoene desaturase (*CrtI*) (Figure 3.2) (Hausmann & Sandmann, 2000, Kanehisa & Goto, 2000, Kanehisa *et al.*, 2016). CrtY catalases the first committed step in production of carotenoids and CrtI is required for the production of many coloured pigments. Therefore it was hoped that using homologous integration to target these genes this would give rise to an albino or colour phenotype which could easily be identified allowing efficient screening of transformants (Niklitschek *et al.*, 2008). A similar system has previously been used to measure the rate of targeted integration in *Aspergillus fumigatus*, targeting the *Abr2* locus yielding a yellow-brown phenotype, rather than the green of wild type *A. fumigatus* (Krappmann *et al.*, 2006).

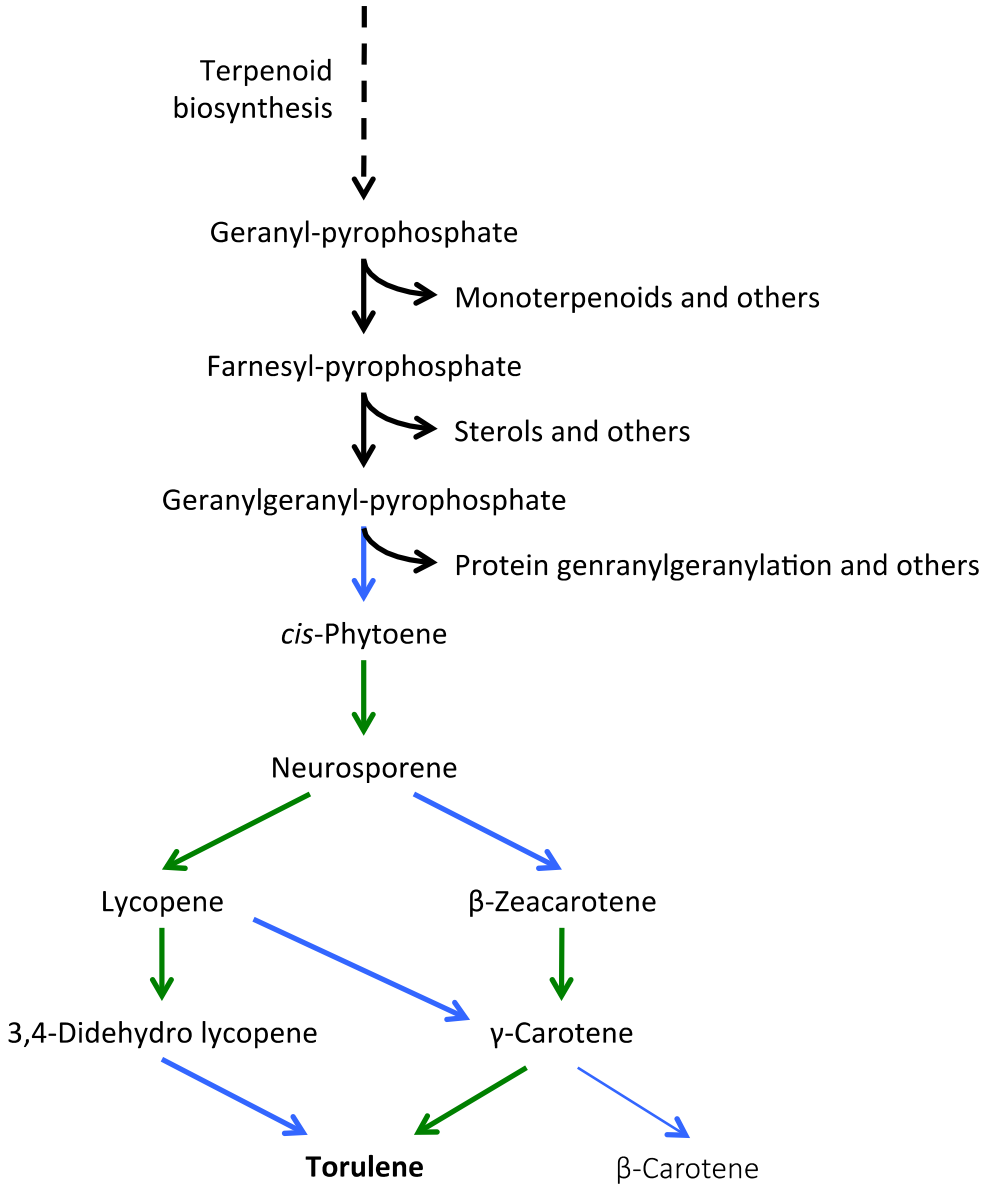


Figure 3.2. *R. toruloides* carotenoid biosynthesis pathway. Reactions catalysed by CrtI are in blue, reactions catalysed by CrtY are in green.

3.1.5 Targeting of Ku genes

As previously noted, T-DNA integration uses the host cell DNA double strand break repair machinery. Non-homologous end joining (NHEJ) is the dominant method of double strand break repair in plants and many fungi (Puchta, 2005, Krappmann, 2007), and is responsible for random integration of T-DNA (Gorbunova & Levy, 1997). The Ku heterodimer is responsible for recognition of the broken DNA ends during NHEJ and acts as the scaffold for recruitment of further factors (Critchlow & Jackson, 1998, Walker *et al.*, 2001). Disruption of one or both Ku genes blocks NHEJ, therefore DNA double strand break repair (and T-DNA integration) must proceed by homologous recombination. As a result *ku*$^{-}$ strains are used to facilitate targeted integration in other fungi normally recalcitrant for homologous integration (Goins *et al.*, 2006, Carvalho *et al.*, 2010). It was hoped to produce a *ku*$^{-}$ *R. toruloides* strain as this would facilitate efficient homologous recombination in the future.

3.1.6 Aims

The aim of the experiments described in this section was to validate *Agrobacterium tumefaciens*-mediated transformation (ATMT) of *R. toruloides* CBS 14 and develop vectors, protocols and selection markers transformation of this strain, with a view to targeting T-DNA to specific chromosomal loci to facilitate targeted gene disruption/deletion.

3.2 Results and discussion

3.2.1 Antibiotic sensitivity of *R. toruloides* CBS 14

While one marker gene is a minimum for selection of transformants, access to more than one marker gene allows for a single strain to be transformed with multiple constructs, and provides alternatives in cases where one marker may be unsuitable. As well as attempting *R. toruloides* CBS 14 transformation with the hygromycin marker used for transformation of *R. toruloides* CBS 349 it was intended to develop further markers genes for use in *R. toruloides* CBS 14 including aminoglycoside-3'-phosphotransferase and gentamicin-3-acetyltransferase, used to confer resistance to the antibiotics G418 and gentamicin respectively.

In order to test the suitability of the antibiotics hygromycin, G418 and gentamicin to select for transformants, the sensitivity of *R. toruloides* CBS 14 to these antibiotics was measured by the agar dilution method (Wiegand *et al.*, 2008). Antibiotic concentrations tested were in the range 0.5–20 μg mL^{-1}. Minimum inhibitory concentrations were found to be 10 μg mL^{-1} for hygromycin and G418 and 20 μg mL^{-1} for gentamicin (Figure 3.3).

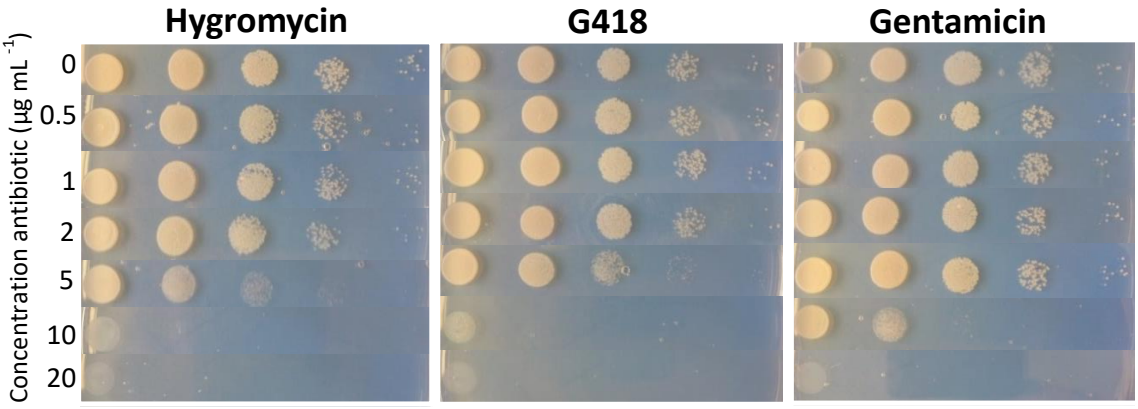


Figure 3.3. Agar dilution assay showing sensitivity of *R. toruloides* to hygromycin, G418 and gentamicin. 10 × serial dilutions of cells were spotted on to YPD agar containing antibiotic at the concentrations indicated and incubated at 30 °C for 24 hours.

For transformation of transformation *R. toruloides* CBS 14, in the case of hygromycin it was decided to initially use 50 µg mL^{-1}, as this was the concentration used for transformation of *R. toruloides* CBS 349. For G418, *R. toruloides* CBS 14 showed a similar sensitivity to that of *S. cerevisiae* (Ernst & Chan, 1985), therefore it was decided to use 150 µg mL^{-1}, the manufacturer's recommended working concentration to be used for selection of transformant *S. cerevisiae*. For gentamicin, the manufacturer's recommended concentration for selection was used (50 µg mL^{-1}).

3.2.2 Isolation of the *R. toruloides* CBS 14 *GPD1* promoter

To drive expression of antibiotic resistance marker genes a suitable promoter is required. For transformation of *R. toruloides* CBS 349 Liu *et al.* (2013) used a 1.43 kb fragment of the homologous *GPD1* promoter to drive expression of the hygromycin resistance marker. In order to transform *R. toruloides* CBS 14, the *GPD1* gene including equivalent promoter fragment was identified in the *R. toruloides* CBS 14 genome and was aligned to the sequence identified by Liu *et al.* (Figure 3.4). It was observed that while the coding region and proximal ~250 bp upstream of the *GPD1* gene is highly conserved, the remainder of the promoter (including the 3' end of the upstream uracil-DNA glycosylase gene) and introns within the *GPD1* gene are divergent. Despite this divergence, it was decided to use the *R. toruloides GPD1* promoter as no other promoters had been characterised for any *Rhodotorula* strain, but to use promoter sequence from strain CBS 14. The equivalent promoter fragment was PCR amplified from the *R. toruloides* CBS 14 genome to be used for construction of plasmid vectors.

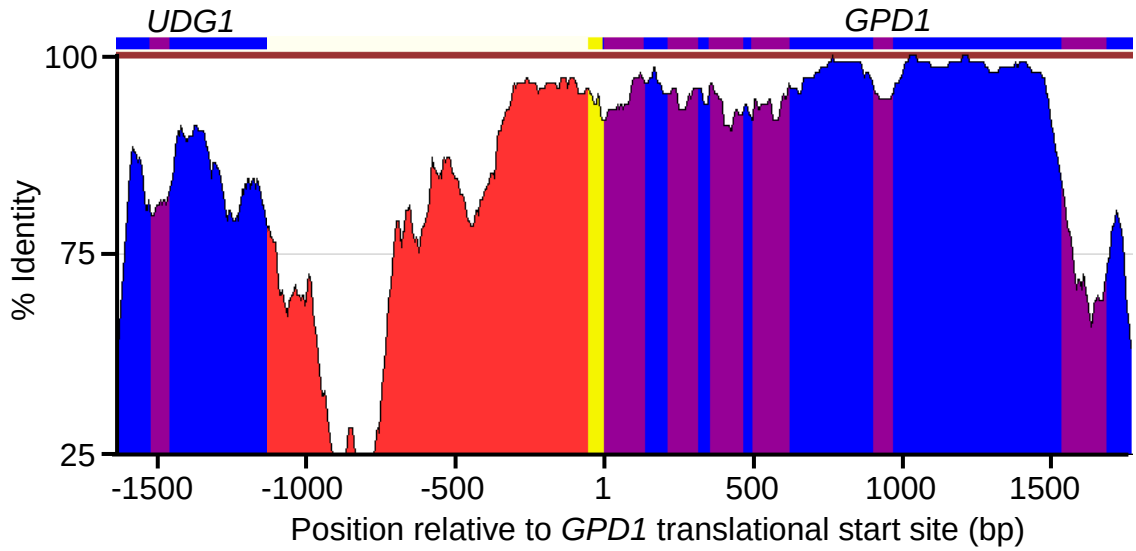


Figure 3.4. Alignment showing the similarity between *R. toruloides* CBS 14 and *R. toruloides* CBS 349 *GPD1* gene and 1.5 kb upstream sequence (including the upstream uracil-DNA glycosylase gene *UDG1*). Intergenic DNA is in red, the 5' UTR is in yellow, coding regions are in blue and introns are in purple. Alignment was performed and % identity plotted using zPicture (Ovcharenko *et al.*, 2004).

The genome for *R. glutinis* strain ATCC 204091 has since been published and is identical to *R. toruloides* CBS 349 (Lin *et al.*, 2014, Paul *et al.*, 2014, Zhang *et al.*, 2016). Overall the genomes of *R. toruloides* CBS 14 and CBS 349 share 87 % homology (Kumar *et al.*, 2012, Zhang *et al.*, 2016), and therefore it would not be unreasonable to consider them not just separate strains but akin to separate species, however these two strains are of different mating types and can mate (Banno, 1967).

3.2.3 *A. tumefaciens*-mediated transformation of *R. toruloides* CBS 14

Three plasmid vectors were constructed for ATMT of *R. toruloides* CBS 14, each with a different antibiotic selection marker. Due to the requirement for codon optimisation of genes for successful expression in *R. toruloides*, synthetic versions of the antibiotic resistance markers were used (Liu *et al.*, 2013). The G418 resistance gene encoding aminoglycoside-3'-phosphotransferase (*APH*(3')) and gentamicin resistance gene encoding gentamicin-(3)-N-acetyltransferase (*AAC*) were synthesised using the most common codon for each amino acid identified in the *R. toruloides* genome; hygromycin phosphotransferase (*HPT*) was synthesised using the sequence of Liu *et al.* (2013) which similarly used the most common codon for any given amino acid except in the case of alanine where CGC was used instead of CGG in two cases out of 37.

The synthetic codon-optimised antibiotic resistance genes were each cloned into the ATMT binary vector pCAMBIA0380 downstream of the PCR-amplified *R. toruloides* CBS 14 *GPD1* promoter, using Gibson assembly. The three marker genes were placed upstream of the nopaline synthase terminator

(T_{NOS}) and downstream of the pUC8 multiple cloning site (MCS) creating plasmids pHyg-Rt, pG418-Rt and pGent-Rt (Figure 3.5 shows pHyg-Rt). Retention of the pUC8 MCS allows future insertion of other genes of interest into the T-DNA portion of this vector.

During domestication of *A. tumefaciens* strain improvements have been made. As a minimum, native T-DNA elements are removed, but also other mutations have been introduced such as disruption of *Rec* genes, in common with other domesticated bacteria (Lazo *et al.*, 1991). Also in common with other domesticated microorganisms, different strains have been developed for use in different situations, for example hypervirulent *A. tumefaciens* strains overexpress virulence genes increasing the efficiency of transformation of plants recalcitrant to ATMT (Lazo *et al.*, 1991). For transformation of *R. toruloides* CBS 349 the hypervirulent *A. tumefaciens* strain AGL-1 was used (Liu *et al.*, 2013). However, strain AGL-1 was unable to transform the basidiomycete *Hypsizygus marmoreus*, whereas other strains including GV3101 were able to efficiently transform this mushroom (Zhang *et al.*, 2014). In order to maximise the probability of transformation of *R. toruloides* CBS 14 two different *A. tumefaciens* strains were tested: AGL-1 and GV3101.

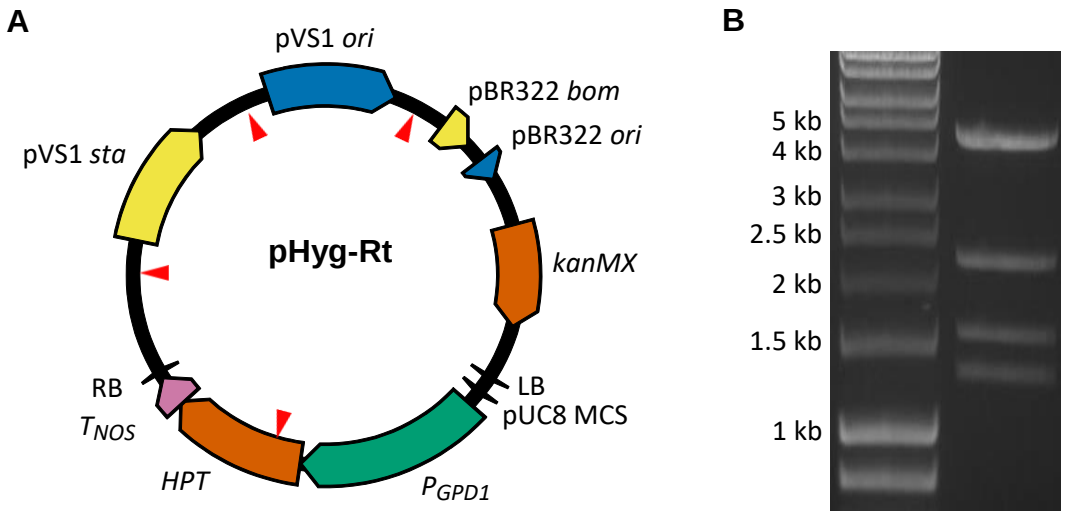


Figure 3.5. pHyg-Rt plasmid vector for transformation of *R. toruloides*. A. Map of pHyg-Rt ATMT plasmid. pCAMBIA0380 was digested with *Pvu*I and a codon-optimised hygromycin phosphotransferase (*HPT*) under the regulation of the *R. toruloides* CBS 14 *GPD1* promoter inserted by Gibson assembly. The T-DNA region is delimited by the left and right borders (LB and RB) and includes the nopaline synthase terminator (T_{NOS}) and the pUC8 multiple cloning site (MCS). Elements in the plasmid backbone are (clockwise) the pVS1 origin of replication and stability region (*sta*) for maintenance in *A. tumefaciens*, the pBR322 basis of mobilisation (*bom*) and origin of replication (*ori*) for bacterial conjugation and maintenance in *E. coli*, and the *kanMX* marker for bacterial selection. With the exception of the *R. touloides* antibiotic resistance marker, pHygRt is the same as pG418-Rt and pGent-Rt. Red arrows indicate *Not*I cut sites. **B. *Not*I digest of pHyg-Rt to confirm correct assembly**. Expected fragment sizes were 4.4 kb, 2.2 kb 1.5 kb and 1.3 kb.

The three plasmids, pHyg-Rt, pG418-Rt and pGent-Rt, were transformed into both *A. tumefaciens* GV3101 and AGL-1 and each used to transform *R. toruloides* haploid strain CBS 14; in each case *R. toruloides* was also mock-treated with untransformed *A. tumefaciens* of the same strain as a negative control. In order to transform *R. toruloides* CBS 14, a two-day culture of *A. tumefaciens* carrying a binary vector with the construct to be transferred was induced to express virulence genes for six hours by dilution into induction medium. 200 µl of this induced *A. tumefaciens*, plus 200 µl of an overnight culture of *R. toruloides* was then spread on to a nitrocellulose membrane placed on solid induction media and incubated at 24 °C for two days. The membrane was then transferred to selective medium containing cefotaxime to kill any remaining *A. tumefaciens* and the appropriate antibiotic to select for transformant *R. toruloides*, and incubated at 30 °C for 2-3 days to allow colonies to develop. In the case of selection with either hygromycin (50 µg mL^{-1}) or G418 (150 µg mL^{-1}) colonies were observed after transformation with plasmids pHyg-Rt or pG418-Rt respectively using both *A. tumefaciens* AGL-1 and GV3101. Also, with these antibiotics, no colonies were observed after mock treatment with untransformed *A. tumefaciens*. However in the case of selection with gentamicin no colonies were observed, including after transformation with pGent-Rt.

In order to compare the efficiency of the different *A. tumefaciens* strains and selection markers for ATMT of *R. toruloides* CBS 14, transformations were performed in triplicate and colonies per plate counted (Table 3.1). No significant difference was observed in the efficiency of transformation using either *A. tumefaciens* strain (ANOVA $F(1,8)=0.12, p=0.74$), either G418 or

hygromycin selection (ANOVA $F(1,8)=0.07, p=0.8$), or any interaction between the two (ANOVA $F(1,8)=0, p=1$).

Table 3.1. Number of colonies per 10 cm plate after transformation of *R. toruloides* with plasmids pHyg-Rt, pG418-Rt and pGent-Rt.

Agrobacterium strain	Antibiotic selection[1]		
	Hygromycin	G418	Gentamicin
GV3101	1870	1670	0
AGL-1	1780	1500	0

[1] Average of three transformations.

In order to confirm the colonies observed were transformants and had not developed antibiotic resistance by other means, colony PCR was performed to confirm the presence of the antibiotic resistance markers. Any remaining *A. tumefaciens* harbouring the plasmid would also cause a positive result in this PCR, therefore a control PCR was performed to amplify the bacterial kanamycin resistance marker from the backbone of the ATMT vector. In most cases the expected result of amplification of G418 or hygromycin resistance marker but not the kanamycin resistance marker was observed (Figure 3.6). In some cases the hygromycin or G418 resistance markers did not amplify, likely due to the limited efficiency of the colony PCR protocol used. Also in three cases the kanamycin resistance fragment amplified, indicating that some contaminating *A. tumefaciens* remained or the kanamycin resistance marker was also transferred to *R. toruloides* (Figure 3.6).

As there was no significant difference in the rate of transformation of *R. toruloides* using each of the *A. tumefaciens* strains, unless noted all further *R. toruloides* transformations were performed using *A. tumefaciens* strain GV3101 as this strain is more robust and easier to manipulate than AGL-1.

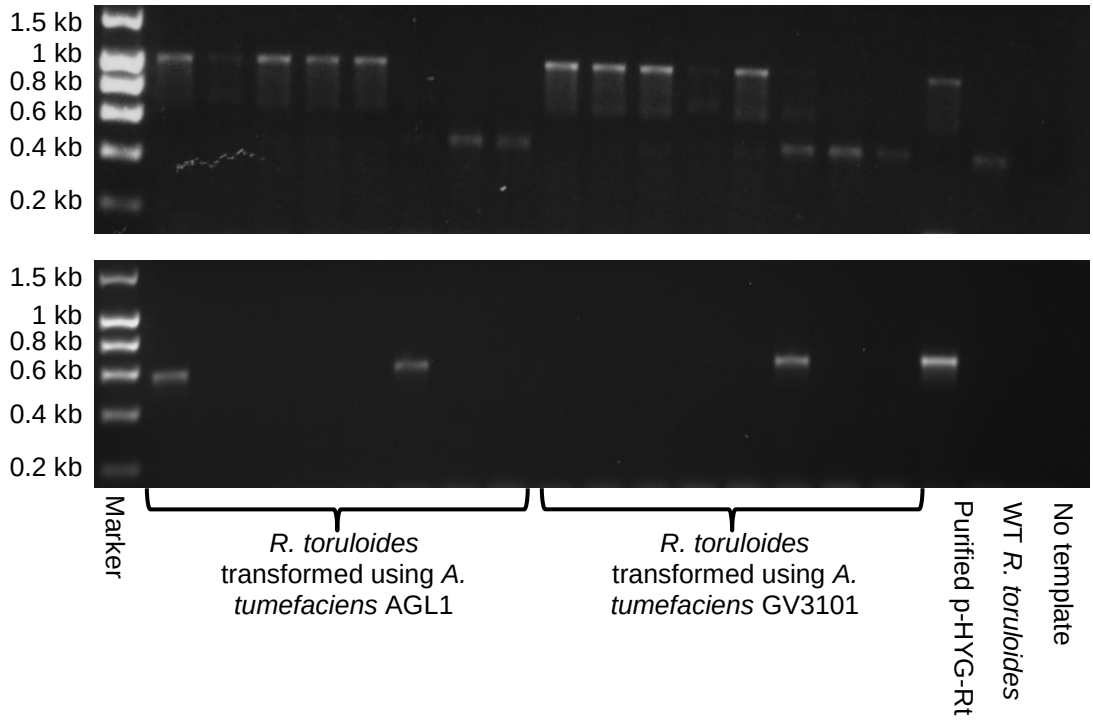


Figure 3.6. *A. tumefaciens*-mediated transformation of *R. toruloides.* Colony PCR was performed to amplify the hygromycin resistance marker from transformant colonies to confirm they are true transformants (upper panel), where a band at 1.1 kb indicates presence of the hygromycin resistance marker. The lower panel shows PCR to amplify the bacterial kanamycin resistance marker, which indicates remaining *Agrobacterium* which could lead to false positives in the upper panel PCR. A band at 0.6 kb indicates the presence of the bacterial kanamycin resistance marker.

Since completion of this work ATMT and selection with the antibiotics hygromycin, bleomycin and nourseothricin has been demonstrated with *R. toruloides* NP 11 (which is almost identical to strain CBS 14 (Kumar *et al.*, 2012, Zhu *et al.*, 2012, Zhang *et al.*, 2016)) (Liu *et al.*, 2013, Lin *et al.*, 2014). Relative to these other antibiotics G418 is cheaper and less toxic, and therefore the preferred choice for transformation of *R. toruloides*.

3.2.4 Transformation of other *R. toruloides* strains

Due to the significant differences between *R. toruloides* strains, it was tested whether the vectors produced could also be used to transform *R. toruloides* CBS 349 and the diploid type strain CBS 6016 which is a product of conjugation between haploid strains CBS 14 and CBS 349 (Banno, 1967). Using the same protocol used for transformation of *R. toruloides* CBS 14, strains CBS 349 and CBS 6016 were treated with *A. tumefaciens* carrying plasmids pHyg-Rt or pG418-Rt, and mock-treated with untransformed *A. tumefaciens*. Colonies were observed on selective media after transformation with both plasmids but no colonies were present after the mock treatment. Resistant colonies were grown overnight in liquid media and spot plated on to media supplemented with hygromycin or (50 μg mL^{-1}) or G418 (150 μg mL^{-1}) and incubated for three days (Figure 3.7). This demonstrates that the vectors, procedures and protocols used are suitable for transformation of all these *R. toruloides* haploid and diploid strains.

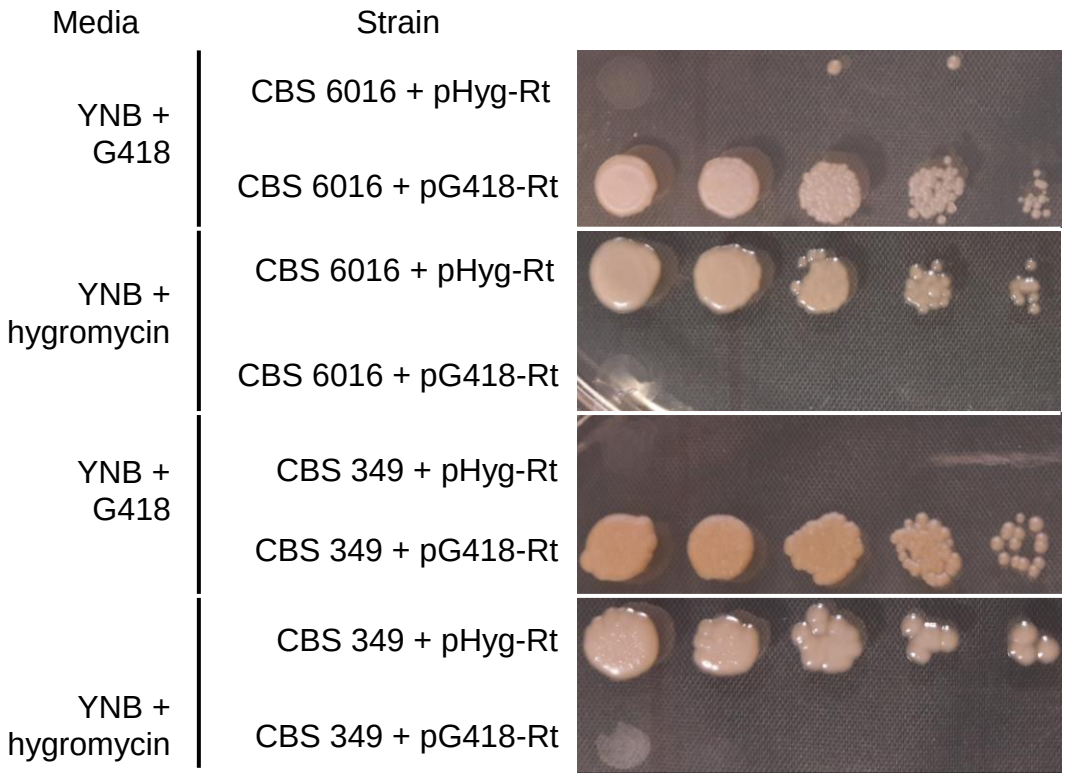


Figure 3.7. Transformation of *R. toruloides* CBS 349 and CBS 6016. *R. toruloides* strains CBS 349 and CBS 6016 were transformed with plasmids pHyg-Rt or pG418-Rt. Transformants were grown overnight in YNB and tenfold dilutions plated on to YNB with either hygromycin or (50 µg mL^{-1}) or G418 (150 µg mL^{-1}). Plates were incubated for three days before being imaged.

3.2.5 Locus of integration of T-DNA

In other fungi and in plant or animal cells ATMT results in insertion of T-DNA at random loci within the genome, sometimes with an apparent bias to expressed regions (Zambryski *et al.*, 1982, de Groot *et al.*, 1998, Kunik *et al.*, 2001, Kim *et al.*, 2007). To test if this is the also case with *R. toruloides*, the locus of integration was determined in each of eight independent transformant clones by thermal asymmetric interlaced PCR (TAIL-PCR), a genome walking technique used to amplify the T-DNA ends and surrounding sequence. This entails three sequential nested PCR reactions performed using specific primers complementary to the known sequence of the T-DNA, in this case approximately 100 bp in from the left border, in combination with a battery of 12 arbitrary degenerate primers designed such that one or more of them should anneal to the surrounding unknown sequence. Due to the GC-rich nature of the *R. toruloides* genome, TAIL-PCR was performed using arbitrary degenerate primers and PCR conditions described by Zhou et al., (2010). These were developed to amplify material from the genomes of Actinobacteria, which have genomic GC-contents between 51 % and 70 % with almost exclusive use of G or C at the wobble position, consistent with *R. toruloides*. After amplification of fragments by TAIL-PCR, Sanger sequencing was performed and compared to the *R. toruloides* CBS 14 genome in order to identify the locus of integration for each transformant.

TAIL-PCR was performed using genomic DNA extracted from 14 independent transformant clones, seven transformed with pHyg-Rt and seven with pG418-Rt. For both plasmids four clones had been transformed using *A. tumefaciens*

AGL-1 and three using *A. tumefaciens* GV3101. From these, the locus of integration was identified in eight clones, three of which were identified using two different non-specific primers. All of these clones had inserts at different genomic locations. Seven had inserts within exons of genes and one in an intergenic region (Table 3.2). Genes disrupted were compared to orthologues from *S. cerevisiae* and *Schizosaccharomyces pombe* to see if these could indicate anything about the biology of *R. toruloides*. In all but one clone genes disrupted are either not essential in both *S. cerevisiae* and *S. pombe* or of unknown function, but in one case insertion was in the *SOG2* gene. *SOG2* is involved in cell polarisation and separation and is essential both *S. cerevisiae* and *S. pombe* (Nelson *et al.*, 2003, Gupta *et al.*, 2013). However this gene is not essential in the basidiomycete yeast *Cryptococcus neoformans* where, although *SOG2* is required for proper cellular polarisation, cells are able to divide and are viable (Walton *et al.*, 2006).

Table 3.2 Locus of integration of T-DNA into the *R. toruloides* CBS 14 genome.

Antibiotic selection	*Agrobacterium* strain used	GACD primer(s)	*S. cerevisiae* orthologue	Insert location	Orthologue essential in *S. cerevisiae*[1]	Orthologue essential in *S. pombe*[1]
Hygromycin	AGL-1	2	NA	Intergenic region between putative nicotinamide N-methyltransferase and a hypothetical protein	NA	NA
		2,3	NA	Within second exon of a hypothetical protein	NA	NA
		1	NA	Within 4th exon of predicted nucleotide binding protein	NA	NA
	GV3101	2	NA	Within coding region of predicted CCCH zinc finger DNA binding protein	NA	NA
		1,3	*NGG1* (transcriptional regulator)	Within 1st exon	No	No
G418	AGL-1	3	*SOG2* (RAM signalling pathway)	Within 4th exon	Yes	Yes
		6		Within 3rd exon of predicted dTDP-4-dehydrorhamnose reductase	NA	NA
	GV3101	1,6	*LUB1* (ubiquitin homeostasis)	Within 11th exon	No	No

[1]Where found; NA indicates no clear orthologue could be identified.

The observed integration of T-DNA into the *R. toruloides* genome is in agreement with observations in other fungi, plants and animals, with T-DNA integrating in an apparently random manner with a possible bias towards expressed regions (Zambryski *et al.*, 1982, de Groot *et al.*, 1998, Kunik *et al.*, 2001, Kim *et al.*, 2007). Also, this is in agreement with Southern blots performed by Liu *et al.* (2013) after transformation of *R. toruloides* CBS 349, which, while not pinpointing the locus of integration, did indicate that in different transformant lines the T-DNA had integrated at different loci; and with the data of Lin *et al.* (2014), who identified the locus of integration in three *R. toruloides* NP 11 transformants by genome walking, two of which were in exons of different genes (a putative serine/threonine kinase and a hypothetical protein) and one upstream of a gene for a hypothetical protein.

3.2.6 In-yeast assembly for construction of vectors for manipulation of *R. toruloides*

In vitro cloning techniques such as restriction cloning or Gibson assembly are usually simple and fast methodologies for assembly of plasmids. However assembly of large or GC-rich fragments often results in a decrease in the efficiency of these techniques. The *R. toruloides* CBS 14 genome has an average GC content of 62 % with some regions being significantly higher. This high GC-content can cause problems during the assembly of plasmids: for example Gibson assembly gave high rates of false positives during assembly of pHyg-Rt and pG418-Rt. For construction of plasmids for targeted genomic integration (see section 3.2.7) restriction cloning and circular

polymerase extension cloning (CPEC) (Quan & Tian, 2014) were also tested, however these also gave low efficiencies and high rates of false positives.

In-yeast assembly uses the simple DNA uptake and efficient homologous recombination machinery of *S. cerevisiae* to assemble multiple overlapping DNA fragments *in vivo* to produce a circular replicating plasmid. Double stranded, linear fragments to be assembled are prepared with approximately 25 bp regions at their ends homologous to the fragments to which they are to be joined and together include an origin of replication and selection marker for transformation of *S. cerevisiae*. These are then co-transformed into *S. cerevisiae*, wherein the yeast homologous recombination machinery joins the fragments together, producing a circular replicating plasmid. This can be extracted by alkaline lysis (Singh & Weil, 2002) and, after passage through *E. coli* to amplify the DNA, used for transformation of *A. tumefaciens* and subsequently *R. toruloides*.

Plasmid pC-G418-YR (Kilaru & Steinberg, 2015) *Zymoseptoria tritici* transformation vector was used as a base for assembly of *R. toruloides* transformation vectors including elements for in-yeast assembly. This plasmid is a derivative of pCAMBIA0380 modified to include an *S. cerevisiae URA3* selection marker and a 2μ origin of replication in the vector backbone. The T-DNA region contained a G418 resistance marker under the regulation of *Z. tritici* α-tubulin promoter and *Neurospora crassa* β-tubulin terminator (Kilaru & Steinberg, 2015). The *Z. tritici* G418 resistance marker and associated promoter were excised and replaced using in-yeast assembly by either the *R. toruloides* codon-optimised G418 or hygromycin resistance markers, each

under regulation of the *R. toruloides* CBS 14 *GPD1* promoter, resulting in production of plasmids pHyg-Rt-YR and pG418-Rt-YR.

After assembly, the junctions formed were Sanger sequenced and diagnostic restriction digests were performed to confirm the overall organisation of each plasmid. The junction sequences confirmed correct insertion of the *R. toruloides* antibiotic resistance markers, but the restriction digest patterns were not as expected. Sanger sequencing of the backbone of plasmid pC-G418-YR revealed a 1.3 kb insertion between the pBR322 origin of replication and the *kanMX* cassette. A BLASTN search identified this sequence as an IS 10 transposable element. As this element is in the backbone of the plasmid it would not be transferred during ATMT and, as it was not expected to affect the efficacy of vectors, it was decided to continue using these plasmids.

To facilitate later experiments a codon-optimised EGFP gene under regulation of the *R. toruloides* CBS 14 *PGK1* promoter (Lin *et al.*, 2014) (mutated to include a *Pml*I cut site at the -7 to -12 position) and the CMV35S terminator, and also incorporating an *Afl*II cut site upstream of the promoter and a *Spe*I site immediately downstream of the EGFP was synthesised. This synthetic construct was inserted into the T-DNA region of both pG418-Rt-YR and pHyg-Rt-YR by in-yeast assembly, creating plasmids pEGFP-Rt-YR-G418 and pEGFP-Rt-YR-Hyg respectively (Figure 3.8A).

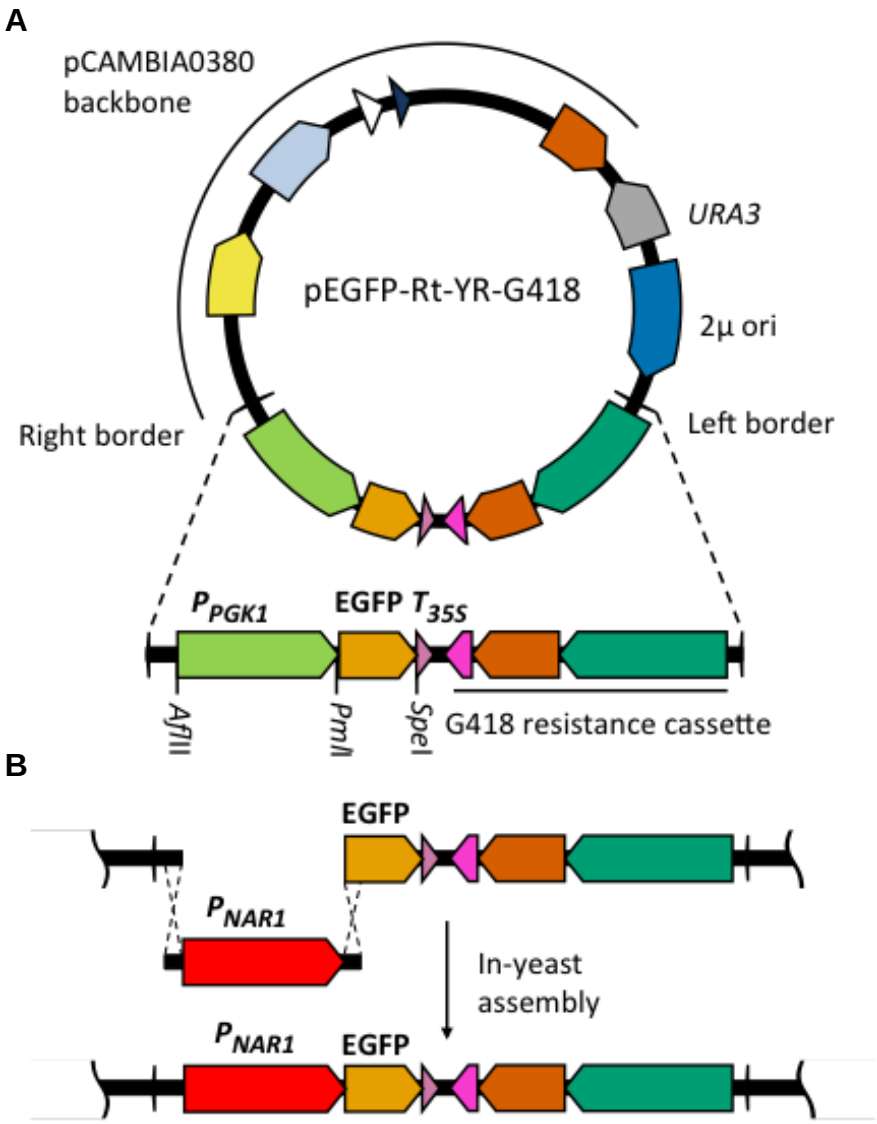


Figure 3.8. Vector and strategy for manipulation of plasmids for transformation of *R. toruloides* by in-yeast assembly. A. *R. toruloides* transformation vector pEGFP-Rt-YR-G418; the T-DNA region to be integrated into the *R. toruloides* genome is shown expanded. The *R. toruloides* G418 resistance cassette consists of a codon-optimised APH(3') gene (orange) under regulation of the *R. toruloides GPD1* promoter (green) and *N. crassa* beta tubulin terminator (pink). The pCAMBIA0380 backbone contains, in a clockwise direction: the right border sequence (RB); pVS1 stability region (yellow) and replication origin (light blue) for maintenance in *A. tumefaciens*; pBR322 *bom* (white) and *ori* (dark blue) for maintenance in *E. coli*; *kanMX* kanamycin resistance cassette (orange). pCAMBIA0380 also provides the left border sequence (LB). Incorporated into the backbone is an *S. cerevisiae URA3* marker and 2µ ARS to facilitate in-yeast assembly. **B.** Cloning strategy for inserting promoters of interest upstream of EGFP gene. The promoter of interest, the *NAR1* promoter in the example shown, is amplified with 25-bp overhangs complementary to regions flanking the insertion site. This is co-transformed into *S. cerevisiae* along with pEGFP-Rt-YR-G418 pre-digested with *Afl*II and *Pml*I. In-yeast homologous recombination inserts the promoter upstream of the EGFP gene in the vector.

Plasmids pEGFP-Rt-YR-G418 and pEGFP-Rt-YR-Hyg are designed such that the *PGK1* promoter or the EGFP gene can easily be exchanged by digestion with *Afl*II/*Pml*I or *Pml*I/*Spe*I respectively and inserting the promoter or gene of interest by in-yeast assembly; an example of promoter replacement is shown in Figure 3.8B. Other cut sites retained in the vector were a *Bgl*II between the *N. crassa* β-tubulin and CMV35S terminators and an *Eco*RI site between the *GPD1* promoter and left border, allowing excision and replacement of the whole EGFP expression cassette by digestion with *Bgl*II/*Afl*II, the antibiotic resistance cassette by digestion with *Eco*RI/*Bgl*II, or whole T-DNA region by digestion with *Eco*RI/*Afl*II.

3.2.7 Targeting of carotenoid biosynthesis genes

As discussed in section 3.2.5, ATMT results in random or quasi-random integration of T-DNA into the host chromosome. Targeting T-DNA to a specific locus would facilitate targeted gene deletion/disruption or, in cases where this is not necessary, it would prevent any unwanted effects resulting from off-target gene disruption or positional effects on transgene expression. In other organisms homologous recombination can be used to target exogenous DNA to predetermined loci and it was hoped to use an equivalent system in *R. toruloides*.

In other basidiomycetes and filamentous fungi gene deletion by homologous integration can be inefficient (Goins *et al.*, 2006, Kuck & Hoff, 2010). In order to maximise throughput when screening for cases of homologous integration, the carotenoid biosynthetic genes *CrtI* and *CrtY* (Figure 3.2) were targeted for

deletion as this should give a strong, visual phenotype of white colonies (Niklitschek *et al.*, 2008).

The target genes were identified in the *R. toruloides* CBS 14 haploid genome by reciprocal BLASTP searches. 1 kb fragments either side of the target genes were amplified by PCR and inserted into pG418-Rt-Yr at the ends of the T-DNA containing the G418 resistance cassette. The resulting plasmids were pCrtI-KO and pCrtY-KO.

Transformations to disrupt *CrtI* and *CrtY* were performed in triplicate giving approximately 300-400 colonies per plate. In order to identify transformants where the T-DNA had integrated at the correct locus, transformed cells were incubated at 30 °C for two days until colonies were visible and then for 2-3 days on the bench at room temperature in the light to allow development of red carotenoid pigments. Unfortunately all colonies observed developed the red colour imparted by carotenoids, indicating that no targeted homologous recombination had occurred (Figure 3.9, Table 3.3).

pCrtI-Ko pCrtY-Ko

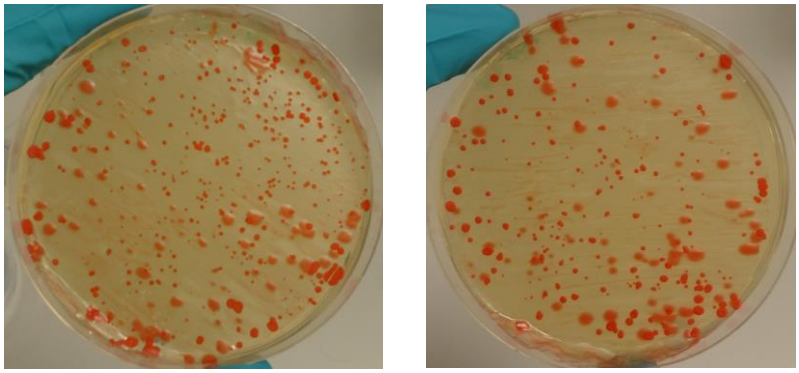


Figure 3.9. Example plates from screen to identify carotenoid biosynthesis mutants. After transformation with plasmids pCrtI-Ko and pCrtY-Ko, plates were incubated until colonies were viable then left on the bench to allow colonies to develop colour. Unfortunately all colonies developed the orange/red colour associated with carotenoid biosynthesis.

3.2.8 Targeting *KU70* and *KU80*

Simultaneously, the genes *KU70* and *KU80* were also targeted for deletion as knocking out one or both of these genes should increase rates of homologous recombination facilitating future gene deletion (Goins *et al.*, 2006, Carvalho *et al.*, 2010, Kuck & Hoff, 2010).

To identify integration at either the *KU70* or *KU80* loci, PCR was performed to amplify across the regions in which homologous recombination was hoped to occur, with one primer complementary to the G418 cassette and one complementary to the sequence outside the 1 kb homology fragment; however in no cases did a fragment amplify, indicating homologous recombination had not occurred (Table 3.3).

In order to increase the rate of homologous recombination, plasmid pKu80-5kbKO was prepared with 5 kb regions homologous to sequences flanking the *KU80* gene (Figure 3.10A,B), and this was transformed into *R. toruloides* CBS 14. Due to the relatively low efficiency of colony PCR when amplifying large fragments from *R. toruloides* an alternative PCR screening strategy was devised that did not involve PCR across the large homology regions. In cases of proper integration the *KU80* gene would be fully excised from the *R. toruloides* genome. In order to screen for integration at the *KU80* locus, transformants were restreaked to single colonies and PCR performed to amplify a 150 bp fragment from the *KU80* gene; in cases of proper integration this fragment would not be expected to amplify (Figure 3.10C). As such a negative result could also result from a failure of the colony PCR, in cases where this initial PCR gave no band, the colony PCR was repeated along with

a second PCR to amplify a 250 bp fragment from the G418 resistance marker to confirm efficacy of the colony PCR. Unfortunately both the *KU80* and G418 resistance gene fragments amplified in every case (Figure 3.10D and Table 3.3).

Table 3.3. Numbers of colonies screened for homologous integration for each of the constructs tested.

Gene Targeted	Homology region size (kb)		Colonies screened
	5'	3'	
CrtI[1]	1	1	~1050
CrtY[1]	1	1	~1050
KU70[2]	1	1	70
KU80[2]	1	1	70
KU80[2]	5	5	300

[1] Screened by red / white selection
[2] Screened by colony PCR

Unfortunately, this study was unable to target T-DNA by homologous recombination to the *CrtI*, *CrtY*, *KU70* or *KU80* loci in the *R. toruloides* CBS 14 genome. This is in accordance with other work since published. Lin *et al.* (2014) attempted targeted integration in *R. toruloides* NP 11 at multiple loci with no success, and Takahashi *et al*. (2014) , attempted targeted integration at the *URA3* locus in the strain *Rhodotorula gracilis* ATCC 26217 (which, from the limited sequence information available is indistinguishable from *R. toruloides* CBS 14), also with no success.

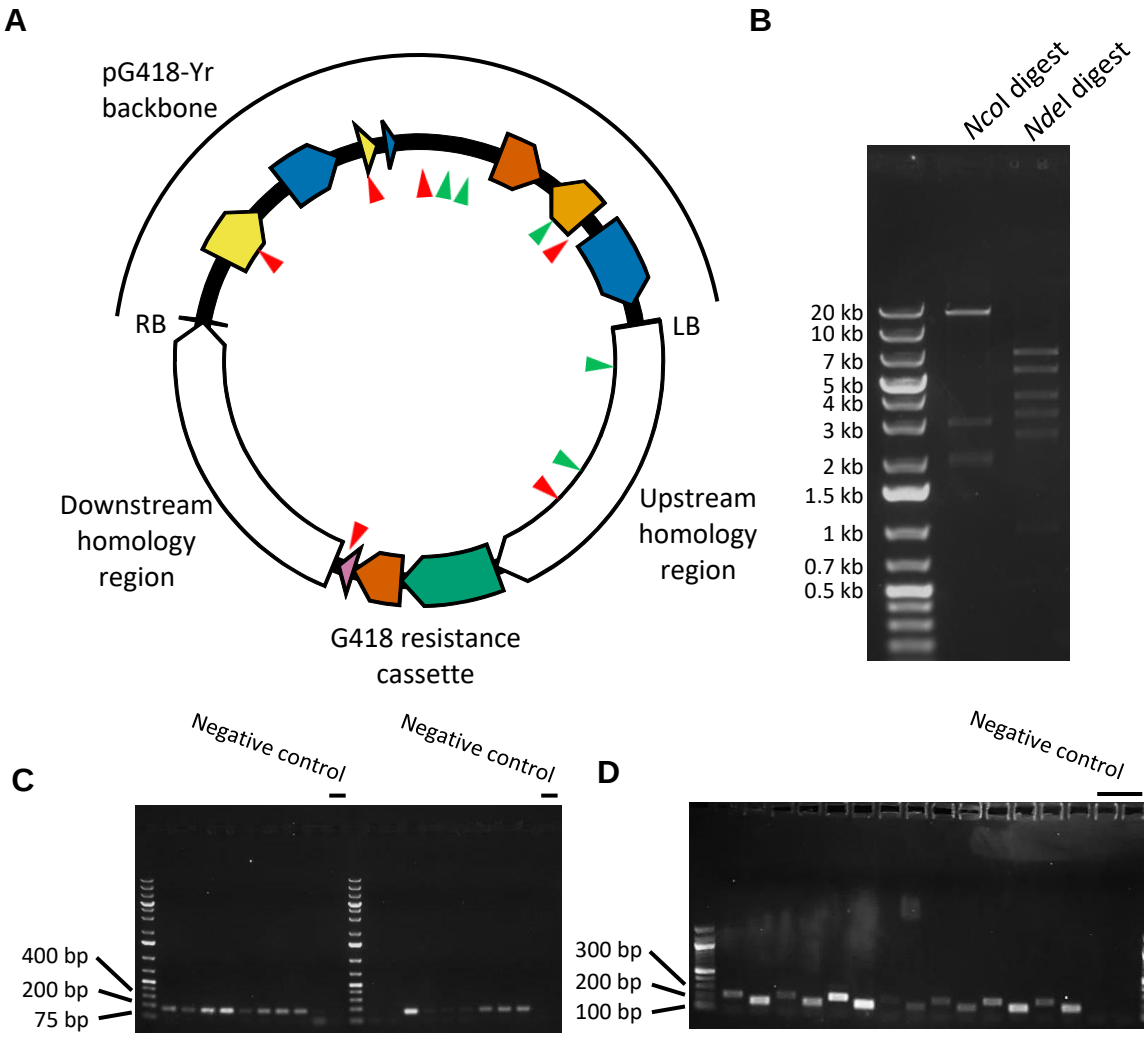


Figure 3.10. Targeted knockout of *KU80* in *R. toruloides*. A. Map of plasmid pKu80-5kbKO. Red arrows indicate *Nde*I cut sites and green arrows indicate *Nco*I cut sites. **B.** Restriction digests of plasmid pKu80-5kbKO with *Nco*I and *Nde*I to confirm identity. *Nco*I digest shows expected bands at 15.6 kb, 2.9 kb, 2.0 kb, 1.9 kb and 0.3 kb (too faint to be seen); *Nde*I digest shows expected bands at 6.8 kb, 5.5 kb, 3.9 kb, 3.2 kb, 2.5 kb and 0.95 kb. **C**. Sample of initial screen of clones transformed with pKU80-5kbKO to identify any clones with *KU80* potentially knocked out. A 150 bp fragment was amplified; where no band was seen clones were taken for further screening. **D**. Second screen to confirm disruption of *KU80* and to confirm efficacy of colony PCR protocol. In each pair of bands, the first (250 bp) is amplification of the G418 cassette as a positive control to confirm the success of the colony PCR, the second (150 bp) is the *KU80* gene fragment.

Since completion of this work, Koh *et al.* (2014) reported targeted integration following ATMT at the *KU70* and *CrtY* loci in strain CBS 349. They used a protocol analogous the one attempted in this work and reported homologous integration at a rate of 30 % with homology regions 1 kb in size at the *CrtY* locus, increasing to 91 % in a *ku70*$^{-}$ strain background. This high rate of homologous recombination relative to the negligible rate observed in this study and by others suggests this may be a result of differences between *R. toruloides* strains CBS 14 and CBS 349.

3.2.9 Carboxin selection for targeted integration

Unfortunately I was unable to identify any cases of homologous integration by transformation and subsequent screening at any of the *CrtI*, *CrtY*, *Ku70* or *Ku80* loci. A second strategy devised was to use homologous recombination to introduce point mutations at known loci, with direct selection of homologous integrants.

The systemic fungicide carboxin is used for control of basidiomycete crop pests including *Ustilago maydis*. After prolonged use, carboxin resistant isolates emerged (Leroux & Berthier, 1988), analysis of which indicated resistance is caused by point mutations in the iron-sulfur containing subunit of succinate dehydrogenase, encoded by the *SDH2* gene. Carboxin acts as a competitive inhibitor of succinate dehydrogenase by binding at the ubiquinone binding pocket, and resistance is caused by point mutations in this pocket preventing binding of the fungicide (Fraaije *et al.*, 2012). The isolated mutant gene was developed as a selectable marker for transformation of *U. maydis*,

and later other fungi including *Z. tritici* (Kojic & Holloman, 2000, Shima *et al.*, 2009).

In order to overcome positional effects on transgene expression after ATMT of *Z. tritici*, Kilaru *et al*. (2015) developed a system using a 5' terminally truncated *SDH2* gene, including the carboxin resistance mutation placed at the end of the T-DNA, upstream of genes of interest to be integrated. If the fragment integrates by homologous recombination at the *SDH2* locus then transformants will be carboxin resistant and transgenes will be at a known locus, however in the case of ectopic integration, only the native *SDH2* would continue to be expressed and transformants would be carboxin sensitive.

It was hoped to use a 5' terminally truncated *R. toruloides SDH2* gene in order to select for homologous integration in *R. toruloides*. Transformation with a full length *SDH2* containing the appropriate mutation, under the regulation of a constitutive promoter, would be expected to confer carboxin resistance to transformed cells, however transformation with a 5'-truncated *SDH2* fragment containing the carboxin resistance mutation would only confer resistance in cases of homologous integration at the native *SDH2* locus (Figure 3.11). This could be used to efficiently select for rare homologous recombination events, and would increase the number of selectable markers available for transformation of *R. toruloides*.

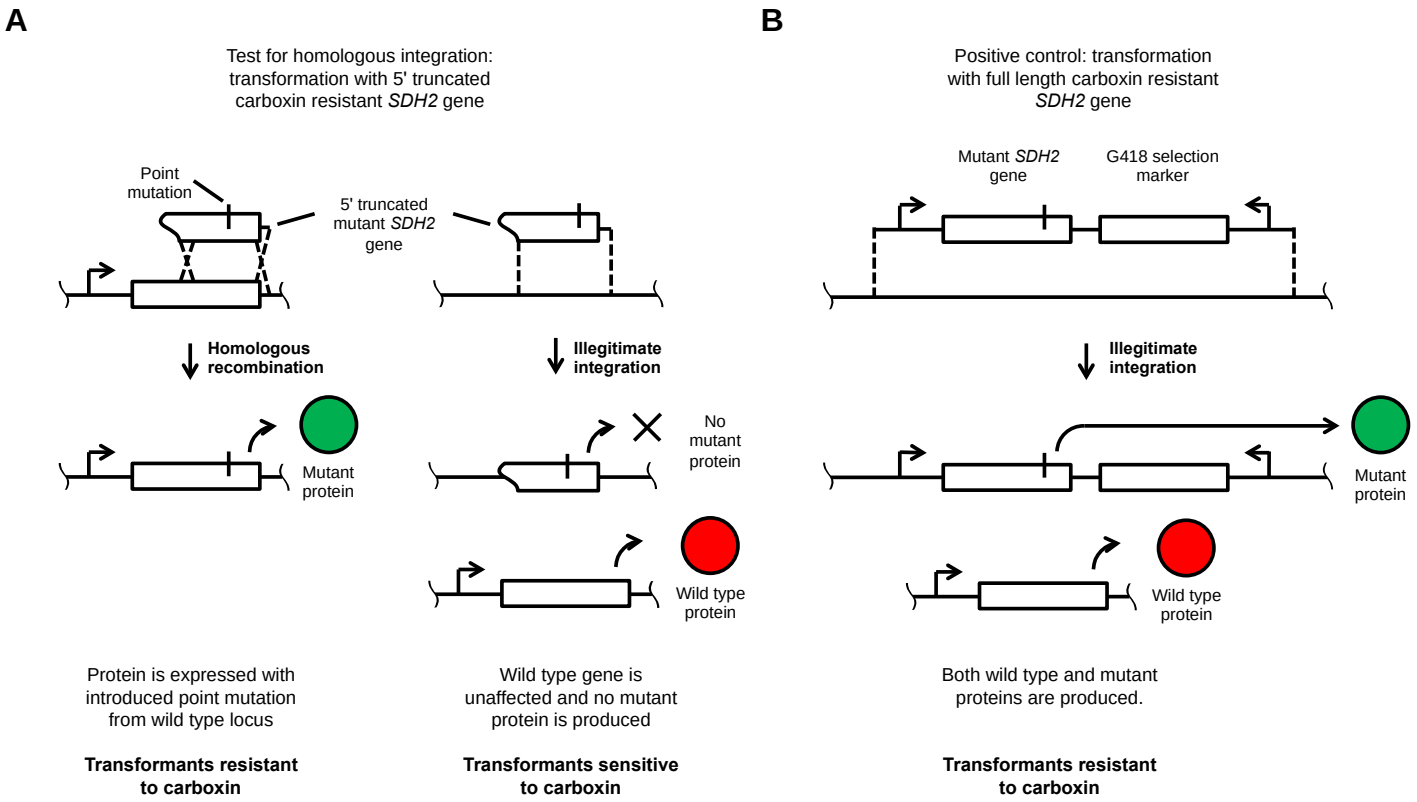


Figure 3.11. Strategy for selection of homologous integrants. A. A 5' truncated version of the *SDH2* gene including the carboxin resistance-conferring mutation is transformed into *R. toruloides*. If this integrates into the host *SDH2* locus by homologous recombination then the resulting strain will be resistant to carboxin, however if this integrates by illegitimate recombination then resulting stains will not be carboxin resistant. An equivalent construct consisting of a truncated *LEU2* gene was used to select for targeted integration at the *R. toruloides* NCYC 1585 *LEU2* locus **B.** As a positive control to confirm the efficacy of the mutant *SDH2* for conferring resistance to the carboxin, full length mutant *SDH2* was also produced and transformed into *R. toruloides* under regulation of the constitutive *PGK1* promoter.

In *Z. tritici* and *U. maydis* a p.H267L (*Z. tritici* numbering) point mutation in the *SDH2* gene causes resistance to carboxin. It was first confirmed that the equivalent histidine residue was conserved in the *R. toruloides* CBS 14 Sdh2 protein. The *SDH2* gene was identified in the *R. toruloides* CBS 14 genome by reciprocal BLASTP and predicted protein sequence was aligned to the equivalent sequences from *Z. tritici* and *U. maydis*. The histidine was identified in a C-terminal region of the protein sequence conserved between the three fungi (Figure 3.12 A).

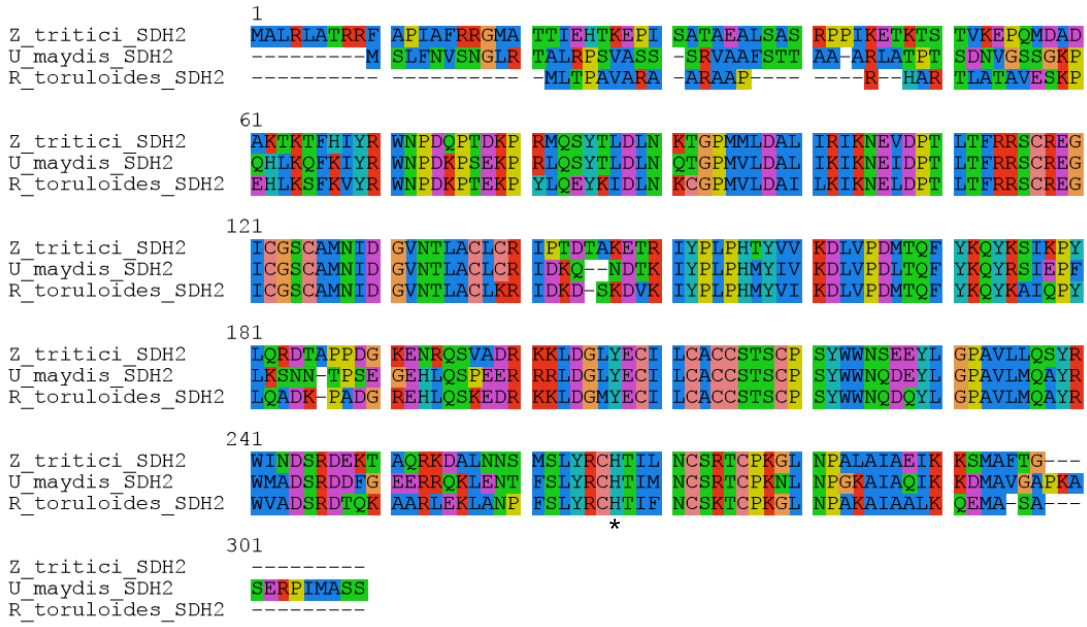
Two constructs were created, one as a positive control to confirm the feasibility of selection with carboxin and a second to screen for targeted integration at the *SDH2* locus. For construction of the positive control a synthetic *SDH2* gene was produced containing a c.671A>T mutation causing a p.H224L substitution (equivalent to the p.H267L mutation in *Z. tritici*) and this was inserted into *Pml*I/*Spe*I-digested pEGFP-Rt-YR-G418 in place of EGFP by in-yeast assembly, creating plasmid pCbx-Rt-YR-G418. To screen for homologous integration a 1 kb fragment was synthesised comprising the 3' end of the *SDH2* gene including the c.671A>T point mutation and downstream intergenic DNA; this was inserted into pEGFP-Rt-YR-G418 digested with *Eco*RI/*Afl*II, replacing the entire T-DNA region between the left and right border sequences, yielding plasmid pCBX-Rt-YR.

In order to confirm the efficacy of the mutant *SDH2* as an *R. toruloides* selection marker, pCbx-Rt-YR-G418 was transformed into *R. toruloides* CBS 14 and transformants were selected using carboxin at concentrations of 0.1, 0.5, 1, 5, 10 and 50 μg mL^{-1} or G418 (150 μg mL^{-1}). After selection with

carboxin at 1 μg mL^{-1}, or greater no colonies were observed, A small number of colonies was observed at a carboxin concentration of 0.5 μg mL^{-1}, however when restreaked to YPD with G418 they failed to grow indicating they were not transformants; with selection at lower concentrations discrete colonies were not observed. Colonies were observed after transformation with pCbx-Rt-YR-G418 and selection with G418.

To test whether the mutant *SDH2* had any effect, clones transformed with pCbx-Rt-YR-G418 and selected for growth on G418 were grown overnight and spot plated on to YPD with 0, 1 and 10 μg mL^{-1} carboxin (Figure 3.12B). From this it can be seen that *R. toruloides* CBS 14 transformed with pCbx-Rt-YR-G418 was able to grow at approximately 10 × higher concentrations of carboxin than wild type *R. toruloides*, however this was not enough to allow selection of transformants after ATMT.

As succinate dehydrogenase is a four-polypeptide complex it was hoped in cases of homologous integration the resulting resistance to carboxin would be greater than was achieved with the positive control, as in this case not all complexes would contain the carboxin resistant iron-sulfur containing subunit. However when pCbx-Rt was transformed into *R. toruloides* and selected for with carboxin, the results observed were the same as with pCbx-Rt-YR-G418, indicating either the phenotype was not strong enough for selection or there was no homologous integration.



B

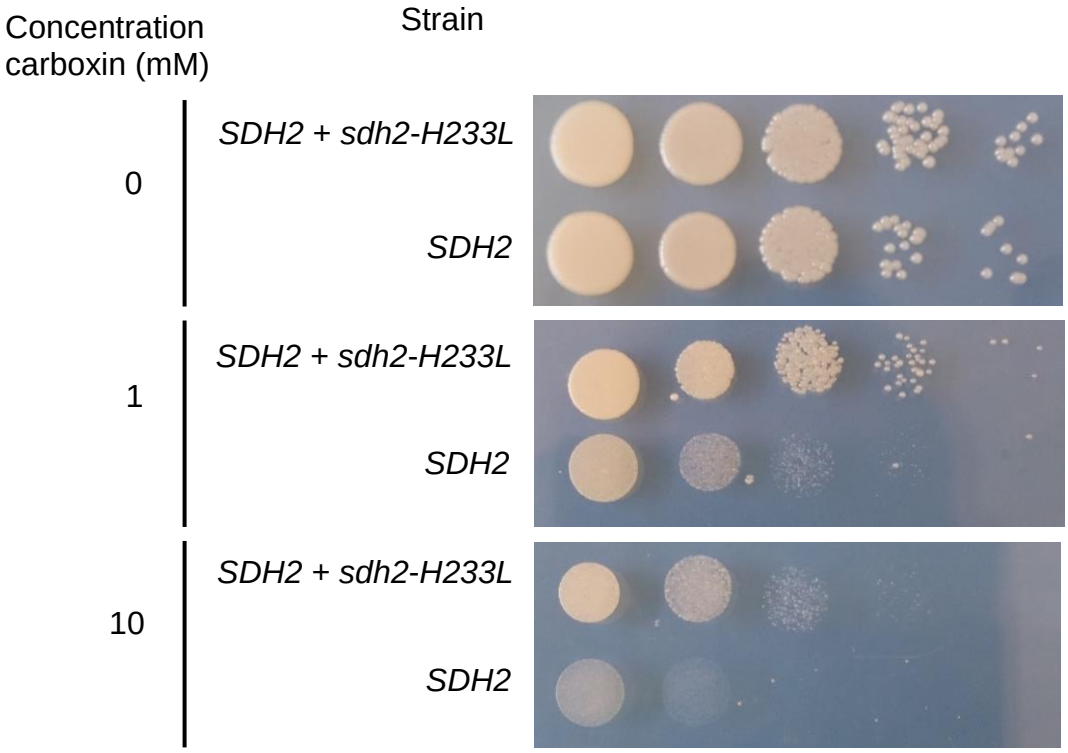


Figure 3.12. Carboxin as a selectable marker in *R. toruloides*. A. Alignment of Sdh2 protein sequences from *U. maydis*, *Z. tritici* and *R. toruloides*. H233 (*R. toruloides* numbering) is indicated by a star. **B.** Growth of *R. toruloides* in the presence of carboxin with and without constitutively expressed *R. toruloides SDH2* p.H233L.

3.2.10 Leucine selection for targeted homologous integration

As carboxin resistance induced by p.H233L point mutation in *SDH2* did not give a sufficiently strong phenotype to select for transformants, it was hoped to identify a point mutation which could give a stronger phenotype. Auxotrophy often results from point mutations in single genes, rectification of which would cause a strong, selectable phenotype. There are no publicly available auxotrophic derivatives of *R. toruloides* CBS 14 and generation of auxotrophic strains would be time consuming, however the *R. toruloides* CBS 349 derived strain NCYC 1585 (Tully, 1985) is publicly available and is auxotrophic for leucine. The *leu2* phenotype is a result of a c.1189C>A mutation in the *LEU2* gene causing a p.G253D substitution in the protein produced (Lin *et al.*, 2012). It was hoped that after transformation with a 5' truncated *R. toruloides* CBS 349 *LEU2* gene including the base affected by the mutation, and selecting for growth in the absence of leucine, transformants would only grow in cases of homologous integration.

It was first confirmed that this strain could be rendered prototrophic for leucine by replacement of the *LEU2* gene. The *LEU2* gene was identified in the *R. toruloides* CBS 14 genome by reciprocal BLASTP, isolated by PCR and inserted into pEGFP-Rt-YR in place of the EGFP gene, creating plasmid pLeu2-Rt-YR-G418. This was transformed into *R. toruloides* NCYC 1585 and transformants selected by growth on YPD with G418 or YNB without leucine. It was noted that a lower rate of transformation was achieved after selection with leucine than with G418 selection. To confirm colonies observed were genuine transformants and not revertants, clones initially selected for with

LEU2 were restreaked to media with G418, on which they were able to grow; the converse was performed and those colonies selected with G418 were able to grow on leucine deficient medium indicating transformation with plasmid pLeu2-Rt-YR-G418 was the cause of reversion of the leucine-deficient phenotype.

To screen for targeted integration two 6 kb fragments were PCR amplified from *R. toruloides* CBS 349; the first of these fragments was a positive control which contained the whole of the *R. toruloides* CBS 349 *LEU2* gene and 4 kb upstream, and the second was a test construct which contained a 5' truncated *LEU2* gene, including 1 kb upstream from the 1189 position mutated in NCYC 1585 and 5 kb downstream (Figure 3.11). These fragments were inserted into *Eco*RI/*Afl*II-digested pEGFP-Rt-YR-G418 by in-yeast assembly to give plasmids pLeu2+ve and pLeu2-test respectively.

During DNA repair by NHEJ the final sealing of the DNA break is performed by DNA ligase IV (Grawunder *et al.*, 1997). The ligase IV inhibitor SRC7 was developed as an adjuvant to increase the effectiveness of cancer treatments which induce double strand breaks in DNA such as radiotherapy or chemotherapy (Srivastava *et al.*, 2012). This chemical has since been applied to increasing the efficiency of genome editing in mammalian cell lines and mouse embryos by blocking illegitimate insertion of exogenous DNA and increasing the relative rate of homologous integration (Maruyama *et al.*, 2015). It was hoped treatment with this chemical would increase the efficiency of homologous integration in *R. toruloides* by blocking ectopic integration.

Plasmids pLeu2+ve and pLeu2-test were transformed into *R. toruloides* NCYC 1585. Co-incubation of *R. toruloides* and *A. tumefaciens* was performed in the presence of 0, 1, or 10 mM SRC7 in triplicate for each condition and colonies counted after recovery on YNB minus leucine. Unfortunately transformation efficiency was low, with only 0, 1 or 2 colonies per plate (Table 3.4); as a result no statistical significance can be assigned to the interaction between concentration of SRC7 and rate of homologous recombination relative to total integration (ANOVA $F(2,8)=3.17, p=0.079$). This low transformation efficiency is likely due to the catch 22 situation that co-incubation with *A. tumefaciens* and selection of transformants was performed in the absence of leucine, which would limit protein production, including of the the Leu2 protein required for leucine biosynthesis. Furthermore additional work would be required to confirm colonies were genuine transformants. Whilst no concrete conclusions can be drawn from this experiment, the rate of homologous integration by the truncated *LEU2* construct at the *LEU2* locus was greater (if not significantly so) in the presence of SRC7, indicating treatment with SRC7 should be further explored as a way of promoting targeted homologous recombination in *R. toruloides*.

Table 3.4. Mean number of colonies per 10 cm plate after transformation of *R. toruloides* NCYC 1585 with 6 kb leucine fragments

Construct	Concentration of SRC7 (mM)		
	0	1	10
Complete *LEU2*	1	0.33	0
Truncated *LEU2*	0.33	0.33	1

[1] Average of three transformations.

3.3 Conclusion

Here I report successful ATMT of *R. toruloides* CBS 14, using dominant selection markers conferring resistance to the antibiotics G418 or hygromycin. The vectors developed are also able to transform the *R. toruloides* haploid strain CBS 349 and the diploid strain CBS 6016, and include cut sites and elements for in-yeast assembly, facilitating rapid assembly of large plasmids for manipulation of *R. toruloides* in spite of the GC-rich DNA of this organism. Furthermore, I was able to demonstrate ATMT using a *LEU2* marker and prototrophic selection of the *leu*$^{-}$ *R. toruloides* strain NCYC 1585. ATMT appears to result in integration of T-DNA at random loci in the genome, in agreement with other studies in *R. toruloides* and other organisms. Unfortunately I was unable to demonstrate targeted integration into the *R. toruloides* CBS 14 chromosome, however this is also in agreement with studies performed in closely related strains.

These plasmid vectors and transformation protocols facilitate molecular genetic manipulation of *R. toruloides*, allowing the development of further tools such as inducible promoters (Chapter 4), and facilitating the expression of other genes of interest, including those for hydrocarbon biosynthesis (Chapter 5).

4 Development of four inducible promoters for use in *R. toruloides*

4.1 Introduction

DNA transformation and transgene expression are the minimum required to genetically modify an organism, however it is often important to regulate transgene expression: for example silencing a toxic gene when it is not required, or switching off an essential gene to study the phenotype. Protein expression can be regulated at the transcriptional (Hu & Davidson, 1987), post-transcriptional (Winkler *et al.*, 2004) and post-translational levels (Mattioni *et al.*, 1994), and examples of all three exist in academia and industry. In yeasts transgene expression is most commonly regulated by inducible promoters, as these do not require production of fusion proteins or the added complexity associated with ribozymes or other RNA regulatory elements.

Several constitutive promoters have been isolated from *R. toruloides* and used to express transgenes, including promoters from the genes *GPD1, FBA1, PGK1, PGI1* and *TPI1* (Liu *et al.*, 2013, Wang *et al.*, 2016). There is however a paucity of inducible promoters characterised for use in *R. toruloides*. The recently isolated *DAO1* promoter is strongly induced when D-amino acids are provided as a carbon source, however it cannot be completely repressed which is a disadvantage for expression of proteins which impede cell growth (Liu *et al.*, 2015b). In addition the D-amino acids required for induction are expensive and may be prohibitive for large,

industrial fermentations. Finally, induction or repression conditions may affect the results of an experiment. Therefore it would be beneficial to have more than one regulatable promoter available for use in *R. toruloides*.

Inducible promoters can either be homologous, where a promoter from the organism of interest is repurposed, or heterologous, where a promoter system from another organism is used. Examples of routinely used homologous inducible promoters include the *S. cerevisiae GAL1* promoter (Johnston & Davis, 1984), or the *S. pombe* nmt1 promoter (Maundrell, 1990). Derivatives of the *E. coli* tetracycline regulatable promoter are commonly used heterologous promoters in yeast, animal (Gossen & Bujard, 1992) and plant cells (Gatz & Quail, 1988). Heterologous promoters have the advantage that they can be engineered to avoid any potential off-target metabolic changes in the organism of interest (Ouyang *et al.*, 2015). However they require expression of response genes, and necessitate the design of chimeric promoters making them difficult to implement, especially given the lack of preexisting knowledge surrounding the structure of promoters in *R. toruloides*. Homologous promoters can be developed by identifying and isolating promoters from native genes which can be regulated by an easily manipulatable stimulus, and placing them upstream of the gene of interest. They do not require expression of response genes or design of chimeric promoters which makes them much faster to develop and have a greater chance of working as expected. Homologous promoters suffer the potential disadvantage that their regulation could potentially cause off target metabolic effects as their use provokes other changes gene expression in the host organism, at the very least from the gene from which the promoter was taken.

This has the potential to bias the results of any experiment using the promoter, however by selecting promoters from pathways orthologous to your gene of interest any off target effects can be minimised. Due to their simpler development and greater likelihood of success it was initially decided to test homologous promoters.

4.1.1 Aims

The aim of this section was to identify and characterise inducible promoters for use in *R. toruloides* CBS 14 in order to provide the means to control expression of introduced genes in different experiments. To this end, it was first necessary to develop fluorescence based reporters for use in *R. toruloides* CBS 14.

4.2 Results and discussion

4.2.1 EGFP and YFP as fluorescent reporters for analysis of gene expression in *R. toruloides*

In order to measure promoter activity a reporter gene was required. pEGFP-Rt-YR-G418 was constructed including a codon-optimised EGFP gene under the regulation of the *R. toruloides* CBS 14 *PGK1* promoter (Section 3.2.6, Figure 3.10). A codon-optimised Venus YFP was also produced and inserted into *Pml*I/*Spe*I-digested pEGFP-Rt-YR-G418 by in-yeast assembly creating plasmid pYFP-Rt-YR-G418.

It was first confirmed that the marker genes are expressed in *R. toruloides* CBS 14. Plasmids pEGFP-Rt-YR-G418 and pYFP-Rt-YR-G418 were transformed into *R. toruloides* CBS 14 and three independent clones grown

over night in YNB, along with two untransformed controls. Cells were imaged microscopically under bright-field and fluorescence microscopy conditions and fluorescence quantified from micrographs. All transformant lines showed fluorescence with both EGFP and Venus YFP (Figure 4.1A). For both EGFP (ANOVA $F(2,337)=141.64$, $p<0.0001$) and Venus YFP (ANOVA $F(2,403)=29.52$, $p<0.0001$) fluorescence was not consistent between transformant lines (Figure 4.1B), likely due to positional effects resulting from of integration site of the EGFP gene in the *R. toruloides* genome or potentially multiple integration events.

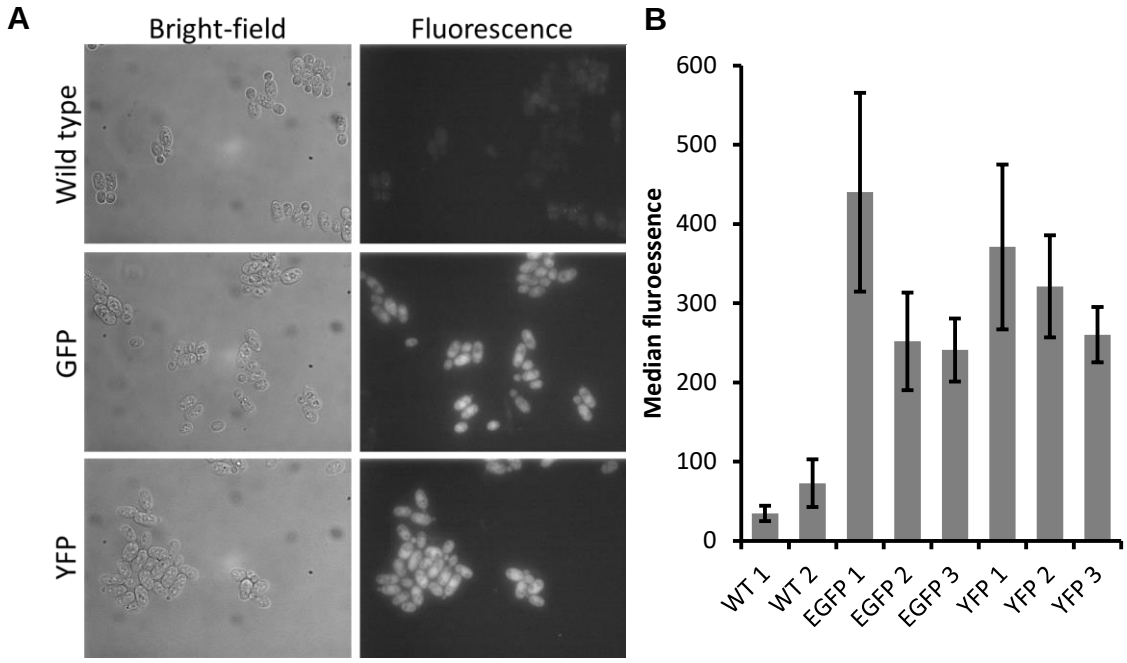


Figure 4.1. Expression of codon-optimised EGFP and Venus YFP in *R. toruloides*. A. Bright field (left panels) and fluorescence (right panels) images of cells untransformed (upper panels), or transformed with pEGFP-Rt-YR-G418 (middle panels) or with pYFP-Rt-YR-G418 (lower panels). **B.** Fluorescence intensity of wild type *R. toruloides* and cells expressing either codon optimised EGFP or Venus YFP. Fluorescence was calculated by measuring the brightness of micrographs and median fluorescence of ≥100 cells from each sample plotted. Error bars indicate median absolute deviation.

Although fluorescence was visible and greatly above background, EGFP expression was not as bright as hoped. In plants (Mascarenhas *et al.*, 1990), animals (Jonsson *et al.*, 1992) and fungi (Lugones *et al.*, 1999) the presence of an intron within the coding region of a gene, proximal to the translational start site, increases gene expression. As the majority of *R. toruloides* genes contain introns (Zhu *et al.*, 2012), it was hypothesised inclusion of an intron may increase expression of EGFP. The introns in the *GPD1* gene have previously been studied and are variously conserved between related strains and species (Liu *et al.*, 2013). Furthermore a *de novo* motif search was performed using MEME (Bailey *et al.*, 2009) on a dataset of 10 000 introns identified by comparison between cDNA and genomic sequence to confirm the consensus sequences at the ends of introns (Zhu *et al.*, 2012) (Figure 4.2A). The sequences of the three promoter-proximal introns from the *R. toruloides* CBS 14 *GPD1* gene were identified and three synthetic EGFP genes were synthesised, each of which contained one of the three promoter-proximal introns from the *GPD1* gene, at the equivalent position relative to the start codon (Figure 4.2B). Also a fourth EGFP gene was synthesised with all three introns. These intron-containing EGFP genes were inserted into *Pml*I/*Spe*I digested pEGFP-Rt-YR-G418 in place of the existing EGFP gene.

Measurement of fluorescence by microscopy was time consuming, therefore, to increase throughput, cellular fluorescence was measured by flow cytometry and median fluorescence determined (Shapiro, 2003). To minimise positional effects of the locus of integration of the T-DNA into the *R. toruloides* genome, each test was performed on three independently-transformed biological replicates and the mean of the three medians reported (Figure 4.2C).

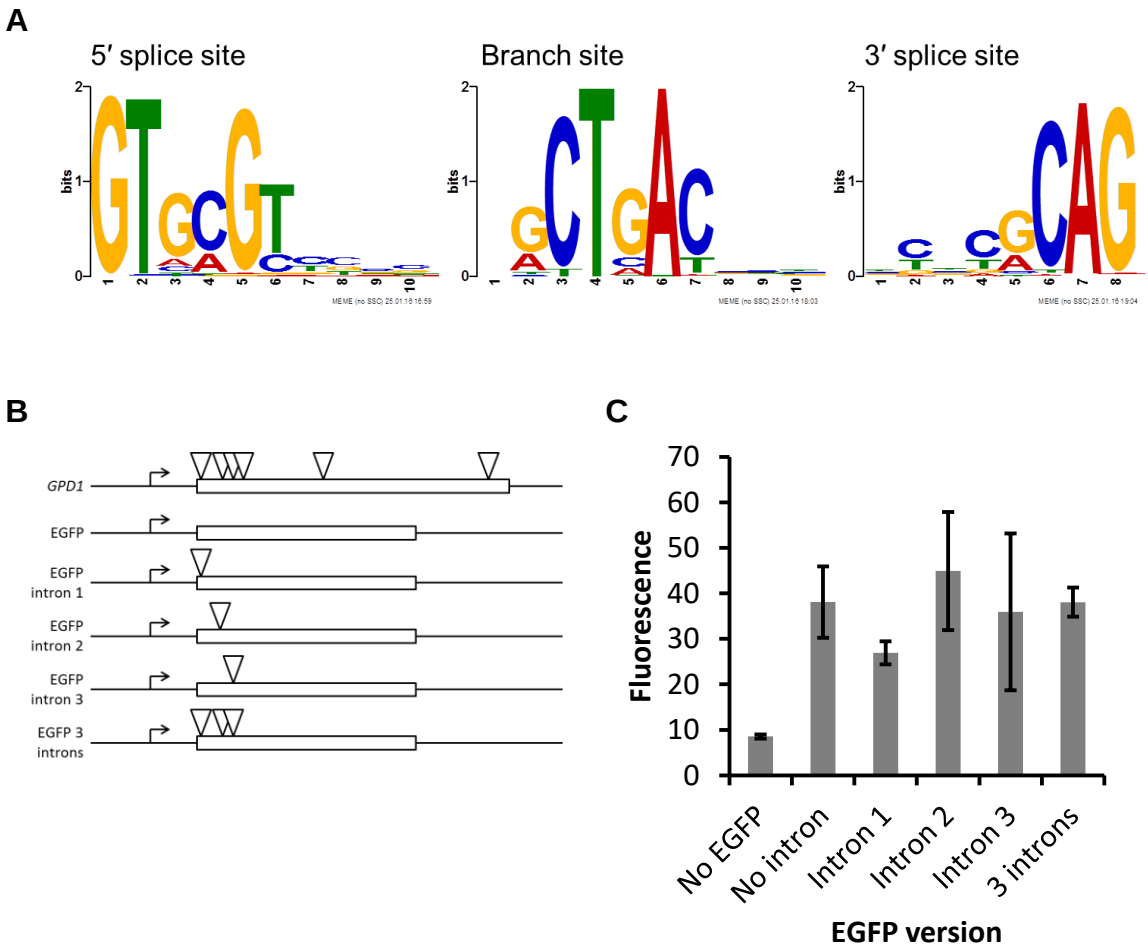


Figure 4.2. Effect of introns on EGFP expression in *R. toruloides*. A. *R. toruloides* consensus 5′ splice site, branch site and 3′ splice sites identified by a *de novo* motif search of 10 000 annotated introns. **B.** Constructs to test the effect of translational start site proximal introns on EGFP expression. Triangles indicate positions of introns. Upper line is the *R. toruloides GPD1* gene, lower lines indicate EGFP with corresponding *GPD1* introns 1, 2 and/or 3. **C.** Fluorescence of *R. toruloides* cells expressing EGFP with *GPD1* translational start site proximal introns. Fluorescence was measured by flow cytometry and the median cellular fluorescence determined for ≥500 000 cells per culture. Tests were performed in triplicate with independent transformants and the mean of the three medians reported. Error bars indicate standard deviation of the three means.

None of the intron-containing constructs resulted in significantly different fluorescence relative to EGFP with no intron (ANOVA $F(4,8)=1.16$, $p=0.38$). Whilst intron-mediated enhancement of the EGFP gene was not observed during this experiment, introns may still effect gene regulation in *R. toruloides* by intron-mediated enhancement or by other means. For example the 3' promoter-proximal intron has been demonstrated to influence regulation the of *R. toruloides* CBS 349 *DAO1* gene (Liu *et al.*, 2015b).

4.2.2 Identification of candidate inducible promoters in *R. toruloides*

To identify a toolset of inducible promoters for use in different situations, potential inducible promoters were screened based on successful use in other fungi. Orthologues of promoters regulated by carbon source, nitrogen source, metabolite availability and copper availability were identified in the *R. toruloides* CBS 14 haploid genome by reciprocal BLASTP hits against their respective genes, and are listed in Table 4.1. *R. toruloides* CBS 14 growth was checked in induction and repression conditions for each candidate promoter. Growth was observed in all media except where galactose was the sole carbon source; as a result *GAL1* and *GAL7* were excluded from further analysis.

Table 4.1. *R. toruloides* candidate inducible promoters.

Gene[1]	Predicted protein	Induced by	Repressed by	Reference
GAL1	Galactokinase	+ galactose - glucose	+ glucose	(Ruff *et al.*, 2009)
GAL7	Galactose-1-phosphate uridyl transferase	+ galactose - glucose	+ glucose	
SGA1	Glucoamylase	+ maltose + starch - glucose	+ xylose + glucose	(Siedenberg *et al.*, 1999) *Aspergillus niger GlaA*
ICL1	Isocitrate lyase 1	+ acetate - glucose	+ glucose	(Barth, 1985)
ICL2	Isocitrate lyase 2	+ acetate - glucose	+ glucose	
NAR1	Nitrate reductase	+ nitrate - ammonium	+ ammonium	(Banks *et al.*, 1993)
THI5	4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase	- thiamine	+ thiamine	(Maundrell, 1990) *Schizosaccharomyces pombe nmt1*
THI4	Thiamine thiazole synthase	- thiamine	+ thiamine	(Manetti *et al.*, 1994) *S. pombe nmt2*
MET16	3′ phosphoadenylsulfate reductase	- methionine	+ methionine	(Solow *et al.*, 2005)
CCC2	Copper efflux pump	+ copper	- copper	(Gebhart *et al.*, 2006) *Histoplasma capsulatum CRP1*
CTR3	High affinity copper transporter	- copper	+ copper	(Labbe & Thiele, 1999)
CTR31	Copper transporter	- copper	+ copper	Paralog of *CTR3*

[1] Gene names reflect *S. cerevisiae* ortholog

4.2.3 GFP screening identifies *NAR1*, *ICL1*, *CTR3* and *MET16* inducible promoters in *R. toruloides*

For each of the 10 potential *R. toruloides* promoters, 1500 bp upstream of the translational start site was amplified by PCR and inserted in place of the *PGK1* promoter, upstream of the EGFP reporter gene in plasmid pEGFP-Rt-YR-G418 (Section 3.2.6, Figure 3.8B)

Each promoter-EGFP construct was transformed into *R. toruloides* haploid strain CBS 14. To identify which candidate promoters can be used as regulatable promoters, cultures were grown for 16 hours under induced and repressed conditions and EGFP fluorescence measured by flow cytometry. To minimise any positional effects from the locus of integration of the T-DNA into the *R. toruloides* genome, each test was performed on three independently transformed biological replicates.

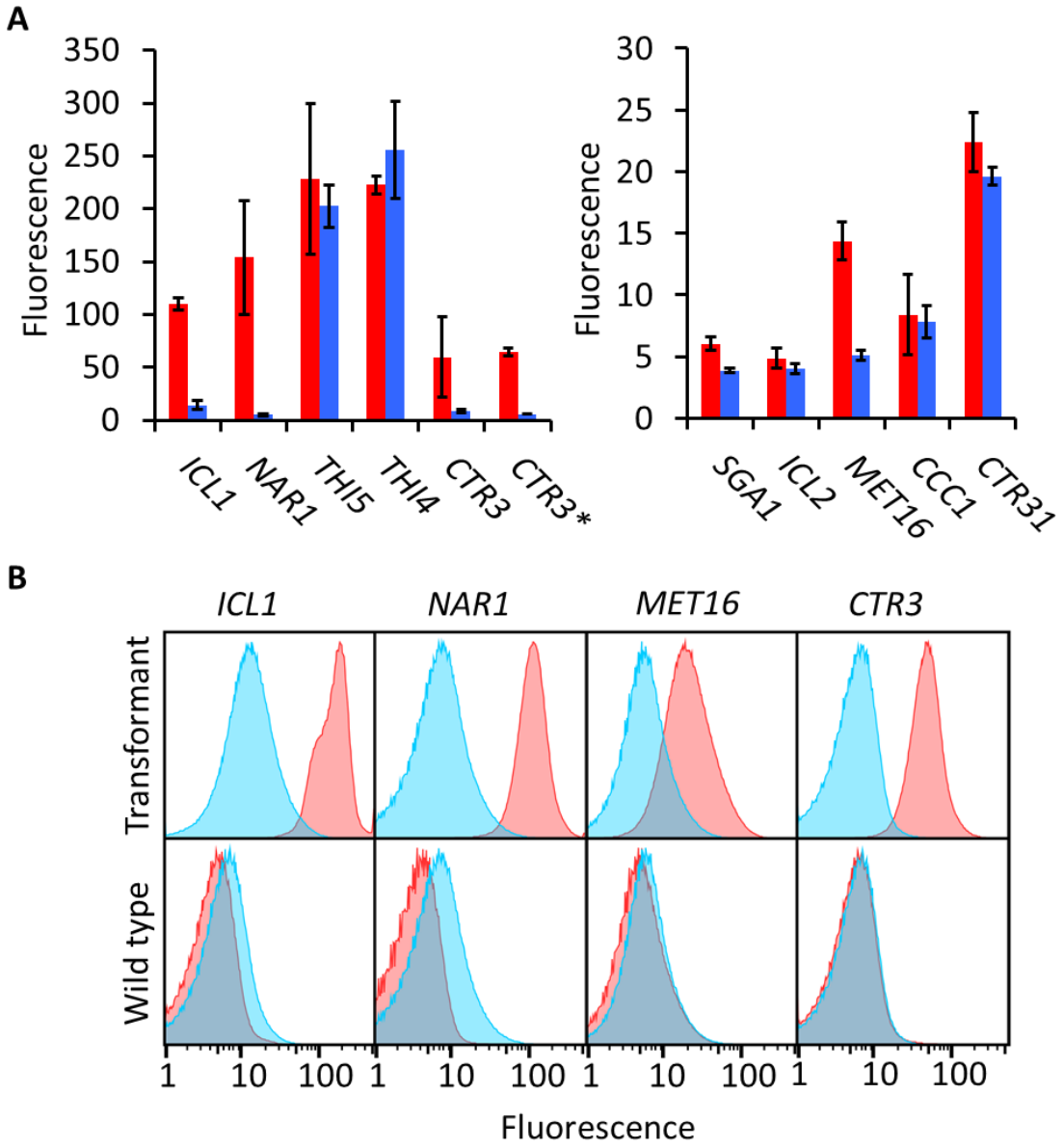


Figure 4.3. EGFP-based screening of ten candidate inducible promoters in *R. toruloides*. (A) Fluorescence of *R. toruloides* cells with EGFP regulated by test promoters after overnight growth in induced conditions (red bars) and repressed conditions (blue bars). Induction and repression conditions are as described in table 4.1 with the exception of *CTR3** where 100 µM BCS was added to induction media. Fluorescence was measured by flow cytometry and the median cellular fluorescence determined for ≥500 000 cells per culture. Bars indicate the mean of three independently transformed biological replicates with the standard deviation shown as error bars. Induction and repression mean values are significantly different ($p < 0.05$) for promoters *ICL1*, *NAR1*, *SGA1*, *MET16* and *CTR3** as determined by student's *t*-test. **(B)** Representative histograms showing fluorescence of cells in induced (red) and repressed (blue) conditions for the *ICL1*, *NAR1*, *MET16* and *CTR3* promoters. Upper panels show transformant cells with EGFP under the regulation of each of indicated promoter; lower panels show autofluorescence of untransformed cells under growth conditions identical to the transformants above.

Of the candidates screened, the promoters of *ICL1*, *NAR1* and *MET16* demonstrated inducibility (Figure 4.3A). The *NAR1* promoter displayed high levels of induced expression surpassed only by the *THI5* and *THI4* constitutive promoters. This promoter also exhibited low expression when repressed (measured induction ratio = 29).

The *ICL1* promoter also displayed high levels of induced expression, however *ICL1* repression was incomplete in the presence of glucose (measured induction ratio = 7.6). This is consistent with activity observed in the oleaginous ascomycete yeast *Yarrowia lipolytica* as well as the economically important *Komagataella* (formerly *Pichia*) *pastoris* (Barth, 1985; Menendez et al., 2003). Acetic acid has been proposed as a feedstock for industrial growth of *R. toruloides* due to its low cost (Huang et al., 2016), and under these conditions the *ICL1* promoter would be induced. Such a system has been proposed for protein production in *K. pastoris*, as an alternative to the commonly used methanol-induced *AOX* promoter (Menendez et al., 2003).

The *MET16* promoter had a low induced expression level (about one tenth the strength of the induced *NAR1* promoter) and a low measured induction ratio. However, under repressed conditions the measured fluorescence was comparable to the autofluorescence of untransformed cells under identical conditions (Figure 4.3B), therefore the apparent induction ratio of 2.8 should be considered a minimum.

The *CTR3* promoter exhibited strong repression in the presence of copper and had a medium level of induction in its absence, however there was a large degree of variation between the replicates. For this reason the copper

chelator bathocuproinedisulfonic acid (BCS) was added to induction medium in all subsequent experiments; this resulted in consistent and significant induction of the *CTR3* promoter (*CTR3** in figure 4.3). The *NAR1* and *ICL1* promoters require changes in nitrogen or carbon sources respectively between induced and repressed conditions; this would have effects on global metabolism whereas the copper starvation conditions for induction of the *CTR3* promoter are unlikely to lead to such gross changes in metabolism (Ouyang et al., 2015). The *CTR3* inducible promoter can therefore be useful where background metabolic considerations are important, such as in a laboratory setting.

Other promoters screened either showed constitutive activity (*THI5*, *THI4*, *CTR31*) or little to no induced fluorescence under the conditions tested (*SGA1*, *ICL2* and *CCC2*).

4.2.4 Gene expression is activated within 4-16 hours of promoter induction

The rate of induction for each of the four promoters was measured by performing a time course over 24 hours from transfer to induction medium, after overnight culture in repression medium (Figure 4.4). Autofluorescence due to carotenoids produced during late log and stationary phase gives high background after 24 hours making measurements unreliable (Kleinegris *et al.*, 2010, Lee *et al.*, 2014).

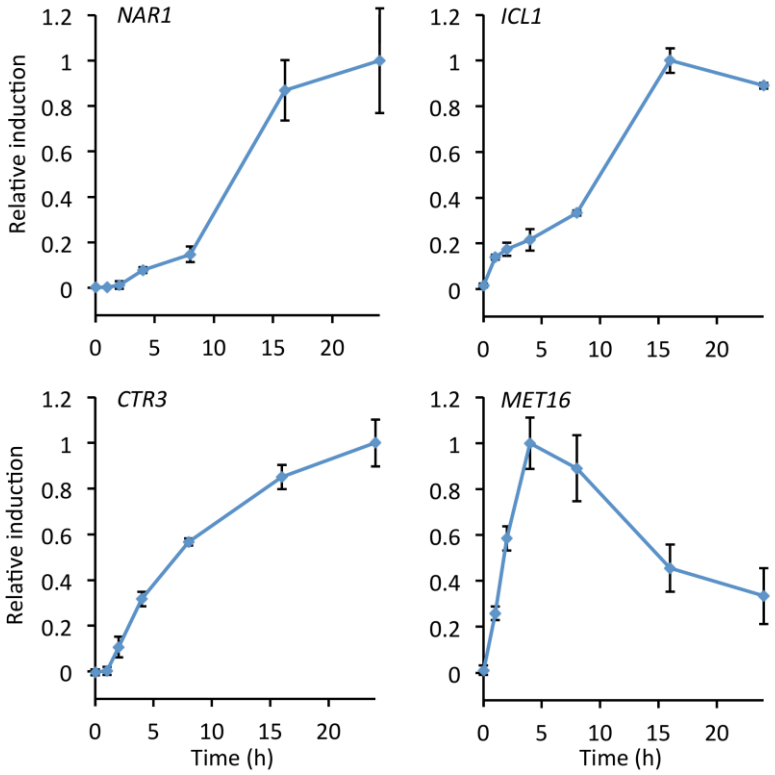


Figure 4.4. Time course showing relative promoter induction up to 24 hours after transfer to inducing conditions. Cultures were grown in repressive medium overnight before cultures were washed, split and transferred to fresh induction/repression medium. Samples were taken at the times indicated and fluorescence measured. Induction was calculated as fluorescence under induced conditions minus fluorescence under repressed conditions and normalised to maximum observed induction. Points show the mean of three independently transformed biological replicates; error bars indicate standard deviation.

The *MET16* promoter was the fastest to induce, reaching a maximum after 4 hours and declining after 8 hours. This promoter may therefore be suitable for experiments where rapid induction is desirable but high-level expression is not required. Both the *NAR1* and *ICL1* promoters showed greatest increases in expression after 8 hours, reaching maxima at around 16 hours. In the presence of the copper chelator BCS, induction of the *CTR3* promoter started at 2 hours and increased asymptotically up to 24 hours (Figure 4.4).

During these time course experiments, the cell density of cultures was measured at each time point and, growth rate under induced and repressed conditions for each promoter calculated. In the case of the *NAR1* and *MET16* promoters, there was no significant difference in growth rate between induction and repression conditions (t-test, $p>0.5$), with observed doubling times of 114 ± 5 minutes. However, in the case of the *ICL1* and *CTR3* promoters there was a significant difference (t-test, $p<0.5$) between growth rates in induction and repression media. When measuring induction of the *ICL1* promoter, cultures with sodium acetate as the sole carbon source (induction conditions) grew slowly relative to cultures with glucose as the carbon source (repression conditions), with measured doubling times of 170 min with acetate relative to 117 min on glucose. When measuring induction of the *CTR3* promoter the measured doubling time in the presence of 100 μM BCS (induction conditions) was 200 min, relative to 139 min with 20 μM $CuSO_4$ (repression conditions). In a laboratory setting this may be problematic when comparing the biology of cultures in induced and repressed conditions, and if using the *ICL1* promoter in an industrial setting this may cause reduction in yield; however this could be overcome by using a two-stage

fermentation, initially growing with glucose and then switching to growth on acetate.

4.2.5 Conditional mutant rescue using the *NAR1* promoter

To investigate controllable mutant rescue the *R. toruloides leu2* mutant strain NCYC 1585 was used (Tully, 1985, Lin *et al.*, 2012). The EGFP gene in vector pEGFP-Rt-YR-G418 was replaced by *LEU2* from *R. toruloides* CBS 14 to give plasmid pLeu-Rt-YR-G418. This construct rescued *R. toruloides* NCYC 1585 growth on leucine deficient medium; transformants could be selected either by growth on leucine-minus medium or by G418 resistance.

The promoter driving the *LEU2* gene was then exchanged for each of the four inducible promoters and these constructs transformed into *R. toruloides* NCYC 1585, selecting for transformants with G418. Transformant strains were grown overnight in induction media supplemented with leucine and spot plated to solid induction/repression media with or without leucine. After overnight culture in inducing conditions all transformants were able to grow in the absence of leucine, indicating mutant rescue by *LEU2* under the transcriptional control of each of the four inducible promoters. On solid medium under repressive conditions, transformants carrying *LEU2* under the regulation of the *NAR1* promoter were unable to grow (Figure 4.5) demonstrating conditional rescue of *leu2 R. toruloides* using the *NAR1* promoter, and confirming low expression levels under repressive conditions for this promoter. Furthermore, when strains with *LEU2* under the regulation of the *NAR1* promoter were grown overnight in repression media supplemented with leucine and spot plated to repressive media, again no

colonies were observed, but when plated out to induction media recovery was approximately 100-fold less than after overnight culture in induction media, indicating the requirement for a recovery period if using this promoter to switch on expression of an essential gene before selection.

Cells transformed with *LEU2* under regulation of *ICL1*, *CTR3* or *MET16* promoters were able to grow under repressive conditions indicating incomplete repression (Figure 4.6). This could reflect strain differences, as the NCYC 1585 *leu2* strain is a derivative of *R. toruloides* strain CBS 349 which shares only 87 % DNA sequence identity with CBS 14 (Kumar *et al.*, 2012, Zhang *et al.*, 2016), although the two strains can mate (Banno, 1967). The *NAR1*, *ICL1*, *MET16* and *CTR3* genes and their respective 1.5 kb promoter fragments were aligned and % identity plotted (Figure 4.7). It was noted that while the *NAR1* and *ICL1* genes and their respective promoters are relatively well conserved, *CTR3* (and to a lesser extent *MET16*) are only poorly conserved between the two strains, possibly explaining the apparent differential regulation. Alternative explanations are possible, for example regulatory elements within a *LEU2* intron enhancing promoter expression, as in the case of the *DAO1* promoter in strain CBS 349 (Liu *et al.*, 2015b), but this would require further study to explore.

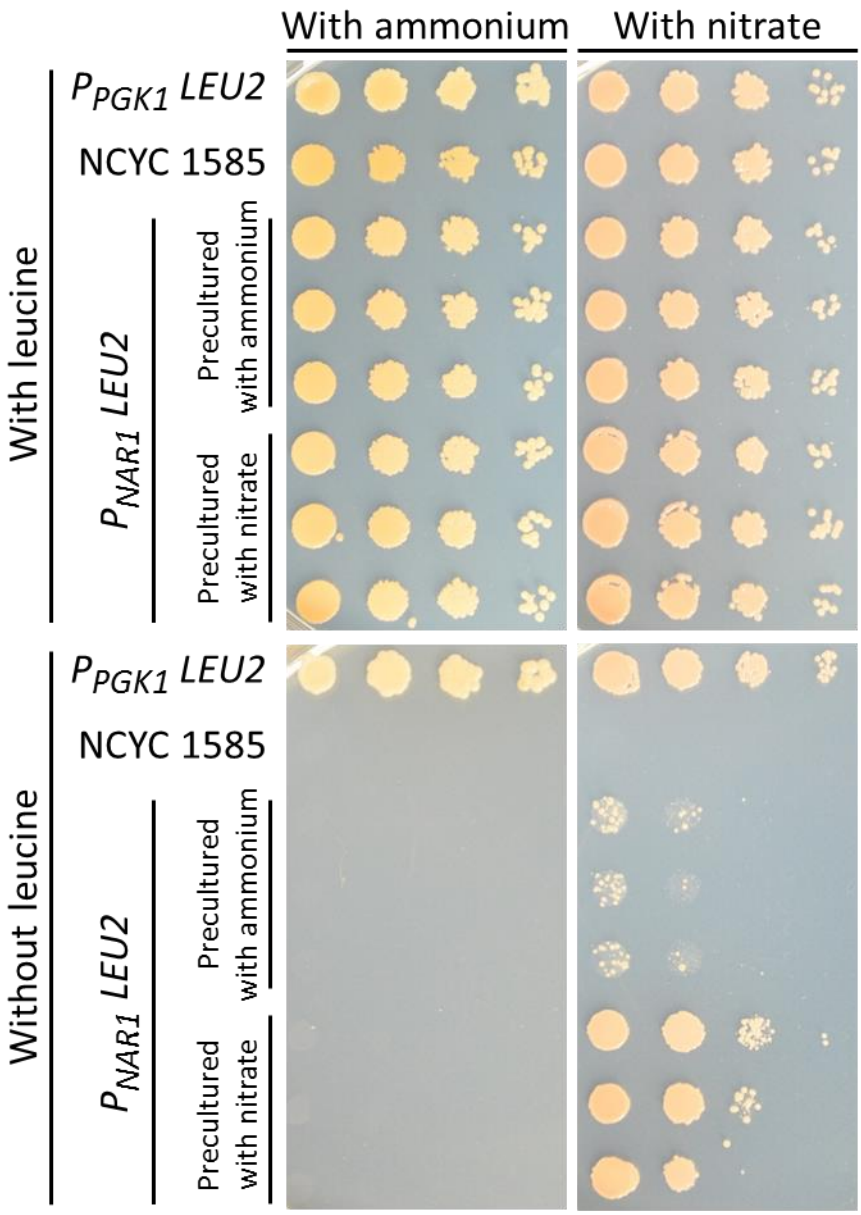


Figure 4.5. Conditional rescue of *leu2* mutant *R. toruloides* strain NCYC 1585 with *LEU2* under regulation of the *NAR1* promoter. Cells from three independent transformant lines were grown overnight in either induction medium (with nitrate) or repression medium (with ammonium), each with leucine (100 mg L^{-1}) and plated on to YNB with 2 % agarose with either 3.5 g L^{-1} ammonium sulfate or 0.78 g L^{-1} potassium nitrate and allowed to grow for 4 days. Cells transformed with *LEU2* under the regulation of the constitutive *PGK1* promoter and untransformed NCYC 1585 cells were included as positive and negative controls respectively.

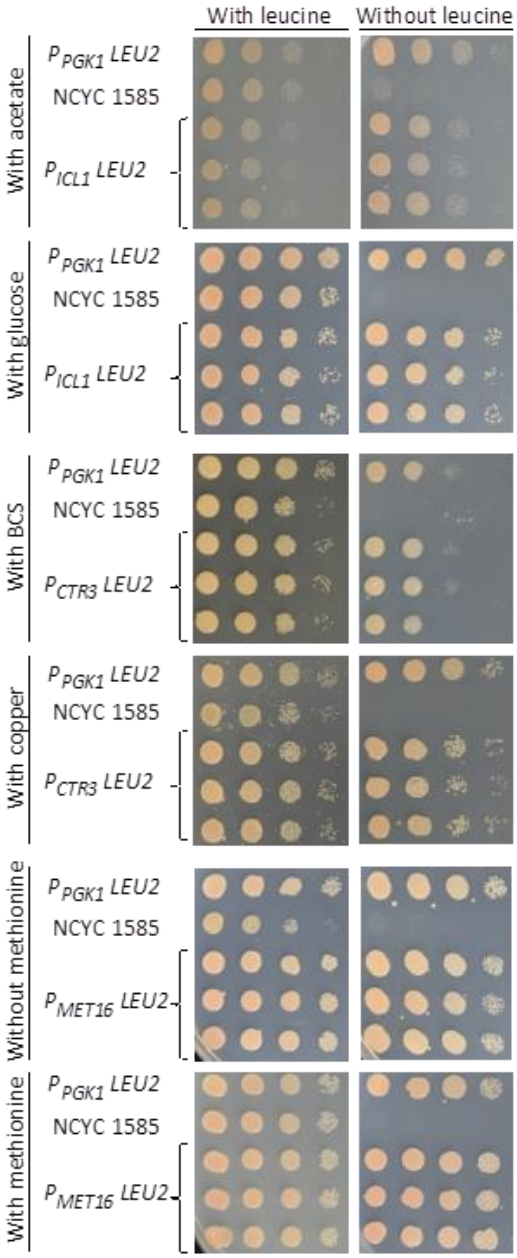


Figure 4.6. Rescue of *leu2 R. toruloides* strain NCYC 1585 with *LEU2* under regulation of the *ICL1*, *CTR3* and *MET16* promoters. In each case three independent transformant lines were grown overnight in induction media with leucine (100 mg L^{-1}) and plated on to solid YNB with 20 g L^{-1} glucose modified for promoter induction or repression, with or without leucine and grown for two days. For induction of P_{ICL1} glucose was replaced with 200 mM sodium acetate; for induction of P_{CTR3}, media were without copper and supplemented with 100 µM BCS, for repression 20 µM CuSO4 was added; for repression of P_{MET16} 1 mM methionine was added. *R. toruloides* NCYC 1585 with *LEU2* under regulation of the constitutive *PGK1* promoter and untransformed *R. toruloides* NCYC 1585 were included as positive and negative controls respectively.

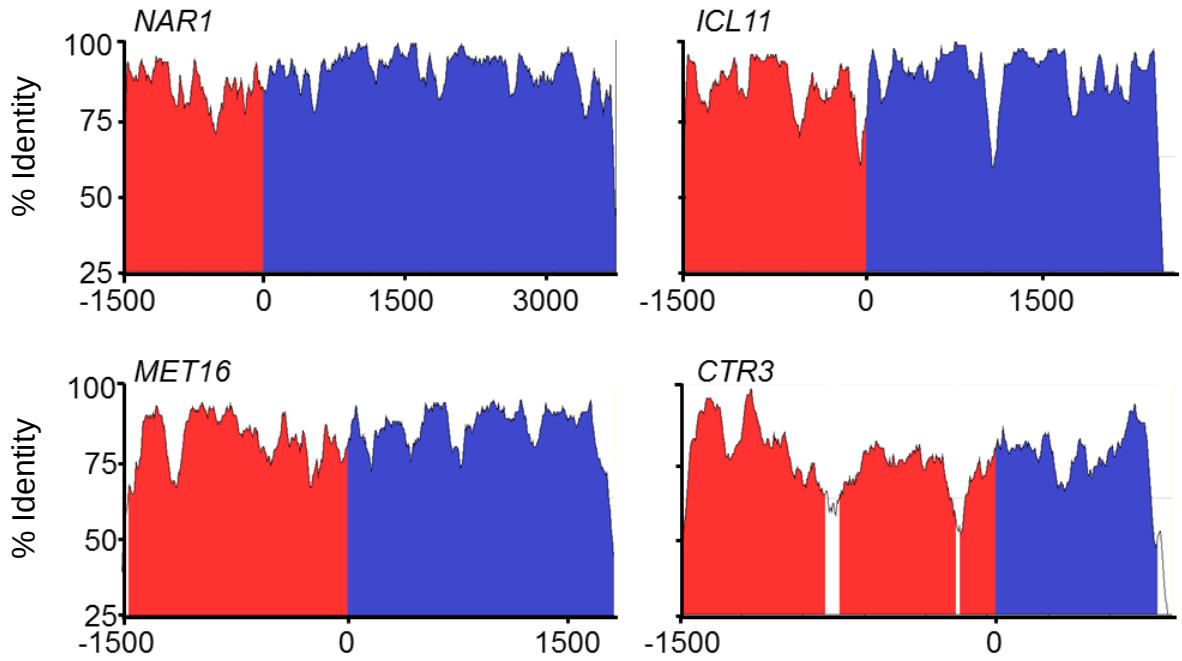


Figure 4.7. Alignment showing the percentage similarity between *R. toruloides* CBS 14 and *R. toruloides* CBS 349 *NAR1*, *ICL1*, *CTR3* and *MET16* genes and promoters. The 1.5 kb promoter fragments are in red and the coding regions (including introns) are in blue. Alignments were performed and % identity plotted using zPicture (Ovcharenko *et al.*, 2004).

4.2.6 Functional dissection of *R. toruloides* inducible promoters

Initially promoter fragments tested were all 1500 bp in length. To identify the minimum size of each promoter required for controllable gene expression and the location of regulatory elements, nested deletions of each of the four inducible promoters were cloned upstream of the EGFP gene (Figure 4.8A) and fluorescence measured for *R. toruloides* CBS 14 transformants under induced and repressed conditions.

With the *NAR1* promoter no activity was observed with the 100-bp fragment but full regulation was observed with fragments 200 bp and longer (Figure 4.8B), demonstrating all necessary controlling elements are present in this short region. Similarly, for the *ICL1* and *MET16* promoters, little or no activity was observed with the 100-bp fragments, full regulation required 400-bp fragments, with 200 bp giving partial activity under induced conditions for *ICL1* (Figure 4.8C,E). *CTR3* promoter cut-downs showed a more interesting pattern: 100- and 200-bp fragments showed little activity, the 400-bp fragment was constitutively active, and the 800-bp and 1500-bp fragments exhibited full regulation (Figure 4.8D). Identification of minimal functional units for these promoters reduces the required size of regulatable gene expression vectors, simplifying cloning of plasmids or the cost associated with synthesis of synthetic DNAs.

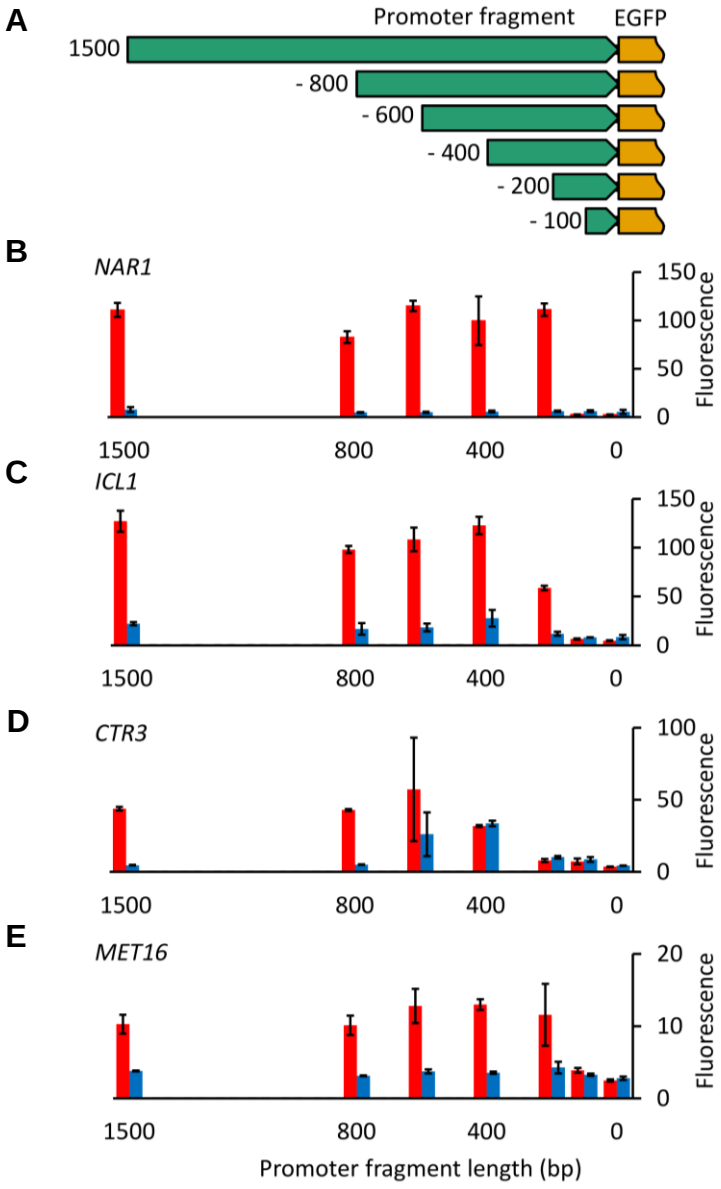


Figure 4.8. Nested deletion of promoter fragments. (A) Diagram of constructs produced. For each of the four inducible promoters fragments 100, 200, 400, 600, 800 and 1500 bp in length upstream from the ATG were cloned upstream of the codon-optimised EGFP gene. **(B)–(E)** EGFP fluorescence of cells carrying truncated versions of *NAR1* **(B)**, *ICL1* **(C)**, *CTR3* **(D)** and *MET16* **(E)** promoters under induced (red) and repressed (blue) conditions. Cells were grown overnight and then transferred to fresh induction/repression media. Fluorescence was measured after 16 hours' growth in induction/repression media (8 hours for *MET16*). Bars show the mean fluorescence of three independently transformed biological replicates; error bars indicate standard deviation. Induction and repression mean values are significantly different ($p < 0.05$) for the following promoter fragment lengths: *NAR1* ≥ 200 bp; ICL1 ≥ 200 bp; *CTR3* ≥ 800 bp; *MET16* ≥ 400 bp and 100 bp.

In this work three biological replicates were performed for each construct or condition and in some cases a large variation was observed between samples. This could be the result of true biological differences between transformant clones, or an experimental artefact, for example if different clones reached a different growth phase (Figure 4.4). Were further work undertaken, in order to confirm which situation is the case it would be appropriate to perform multiple experimental replicates of each biological replicate; comparing the variation between biological and experimental replicates would indicate the source of the variation.

Were the variation found to be a genuine biological effect, potential causes could be positional effects resulting from the locus of integration or multiple integration events. The former case is discussed in section 3.1.3 and could be overcome by targeting T-DNA to predetermined loci within the *R. toruloides* genome, however this has proved challenging. In the case of multiple integration events this could be tested for by performing quantitative-PCR, only selecting clones which have been confirmed to be single integrants.

4.2.7 Motifs in *R. toruloides* inducible promoters

To identify functional elements within essential regions of the four inducible promoters, a *de novo* motif search was performed using MEME for conserved elements between orthologous promoters in *R. toruloides* and related members of the Pucciniomycotina (see 2.8.1 for a list of the related species). In both the *ICL1* and *CTR3* promoters, CT-rich boxes were identified in the -50 to -40 region relative to the start codon (Figure 4.9). Similar elements have been observed in the *R. toruloides GPD1* and *DAO1* promoters (Liu *et*

al., 2013, Wang *et al.*, 2016) indicating this is a highly conserved element in *R. toruloides*. Such an element has also been observed in other filamentous fungi where it is proposed to be responsible for targeting the translational start site (Punt *et al.*, 1990).

In the *CTR3* promoter a second conserved box was identified at -583 to -602 (Figure 4.9). The 400-bp promoter fragment showed constitutive induction, the 600-bp fragment exhibited variable repression and the 800-bp fragment full repression in the presence of copper; this sequence element could therefore be responsible for repression of this promoter in the presence of copper. Other instances of this element were identified in *R. toruloides* promoters using FIMO (Grant *et al.*, 2011) and the genes adjacent to the top 10 hits identified. Apart from *CTR3*, the top hit was upstream of a vacuolar ABC heavy metal transporter, a gene likely to be regulated by copper, and the second hit was in the promoter for salicylate hydroxylase, the product of which (catechol) is toxic in the presence of heavy metals (Schweigert *et al.*, 2001) and thus would likely be repressed in the presence of copper. The motif was also identified 283 bp downstream of a second gene annotated as copper transporter. Given the range at which this element acts it is possible that this element may act on the promoter of this gene from this location.

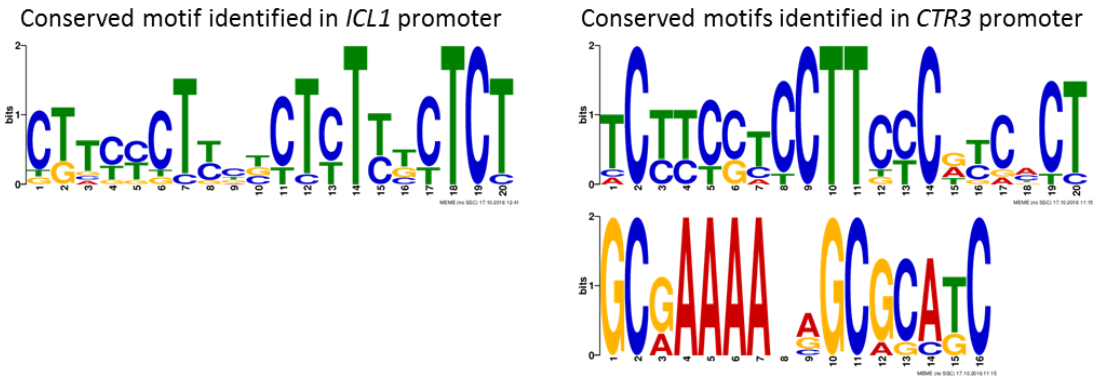


Figure 4.9. Conserved motifs identified in *ICL1* and *CTR3* promoters. Orthologues of identified inducible promoters from other members of the Pucciniomycotina and were subjected to a *de novo* motif search using MEME. Conserved elements were identified in the *ICL1* and *CTR3* promoters and are reported as LOGO plots.

4.2.8 Bioinformatic search for globally conserved promoter elements in *R. toruloides*

With the exception of the CT-rich sequences identified closely upstream of the translational start sites of the *ICL1* and *CTR3* genes, no putative core eukaryotic promoter elements such as TATA or CAAT boxes were observed in the four inducible promoters. In order to guide any further study of promoters in this in *R. toruloides* it was hoped to identify such elements. Practical methods to enrich motifs such as chromatin immunoprecipitation-sequencing could be used, but due to cost and time were beyond the scope of this work, therefore a bioinformatic approach was taken in order to try and identify conserved core promoter motifs.

To identify motifs of associated with a given gene or function, homologous motifs from related species (see section 4.2.7), or co-regulated motifs from one organism can be searched for conserved elements. While core motifs would not be expected to be conserved in all promoters or enriched in any particular subset, they would be expected to be conserved in an appreciable proportion. By searching for motifs enriched within a set of all *R. toruloides* promoters it was hoped to identify conserved core promoter elements. Dozens of algorithms have been developed to identify motifs enriched between sets of biological sequences at potentially variable loci (Hu *et al.*, 2005). MEME (Bailey *et al.*, 2009) is one such algorithm; relative to other comparable algorithms MEME is highly sensitive to detecting different types of motif due to its ability to estimate motif length (Hu *et al.*, 2005), and scales well to the input of datasets containing longer sequences such as the dataset

used in this study. MEME maintains a large user base and continues to be updated (Tanaka *et al.*, 2014, Bailey *et al.*, 2015). Whilst MEME can be configured to identify motifs found any number of times within a given input sequence, due to the extra computational power this requires and reduction in sensitivity this can cause when identifying certain motifs (Bailey *et al*., 2009), MEME was configured as such that it would identify motifs based on one or zero occurrences within a sequence (ignoring further occurrences).

The training set used the *R. toruloides* NP 11 genome, which is almost identical to that of strain CBS 14 (Zhu *et al.*, 2012, Zhang *et al.*, 2016). Zhu *et al*. (2013) annotated this genome using transcriptomic data and identified transcriptional and translational start sites. Using this they published a dataset of 7972 promoters, where 1 kb upstream of each translational start site was extracted, and trimmed where promoter sequences overlapped upstream genes or other promoters. *De novo* motif searches were performed on this dataset using MEME, and the top 10 conserved elements in *R. toruloides* NP 11 promoters reported (Figure 4.10A). Motif 1 and motif 2 appear to be reverse complements of each other, as is also the case with motifs 3 and 4, motifs 6 and 10 and possibly motifs 7 and 8. Therefore, this result represents six unique motifs identified (Figure 4.10A).

In the four studied inducible promoters motif 1 was identified at a locus overlapping the CT-rich element previously identified in the *CTR3* promoter; similarly motif 3 was identified at a locus overlapping the CT-rich motif previously identified in the *ICL1* promoter. Furthermore, motif 1 was identified

in the *MET16* promoter 77 bp from the translational start site, and motif 3 was identified 98 bp from the translational start site of *NAR1*.

Motifs 7 and 8 are approximate reverse compliments of each other and were identified in 949 and 1053 promoters respectively. Although these were not identified in the interspecies motif searches, motifs 7 and 8 were identified at -463 and -368 respectively in the *CTR3* promoter and motif 7 was also identified at position -269 in the *MET16* promoter. In all three cases these motifs were in fragments required for full promoter activity, indicating these motifs are potentially general transcriptional activatory sequences; however this would require more work to confirm.

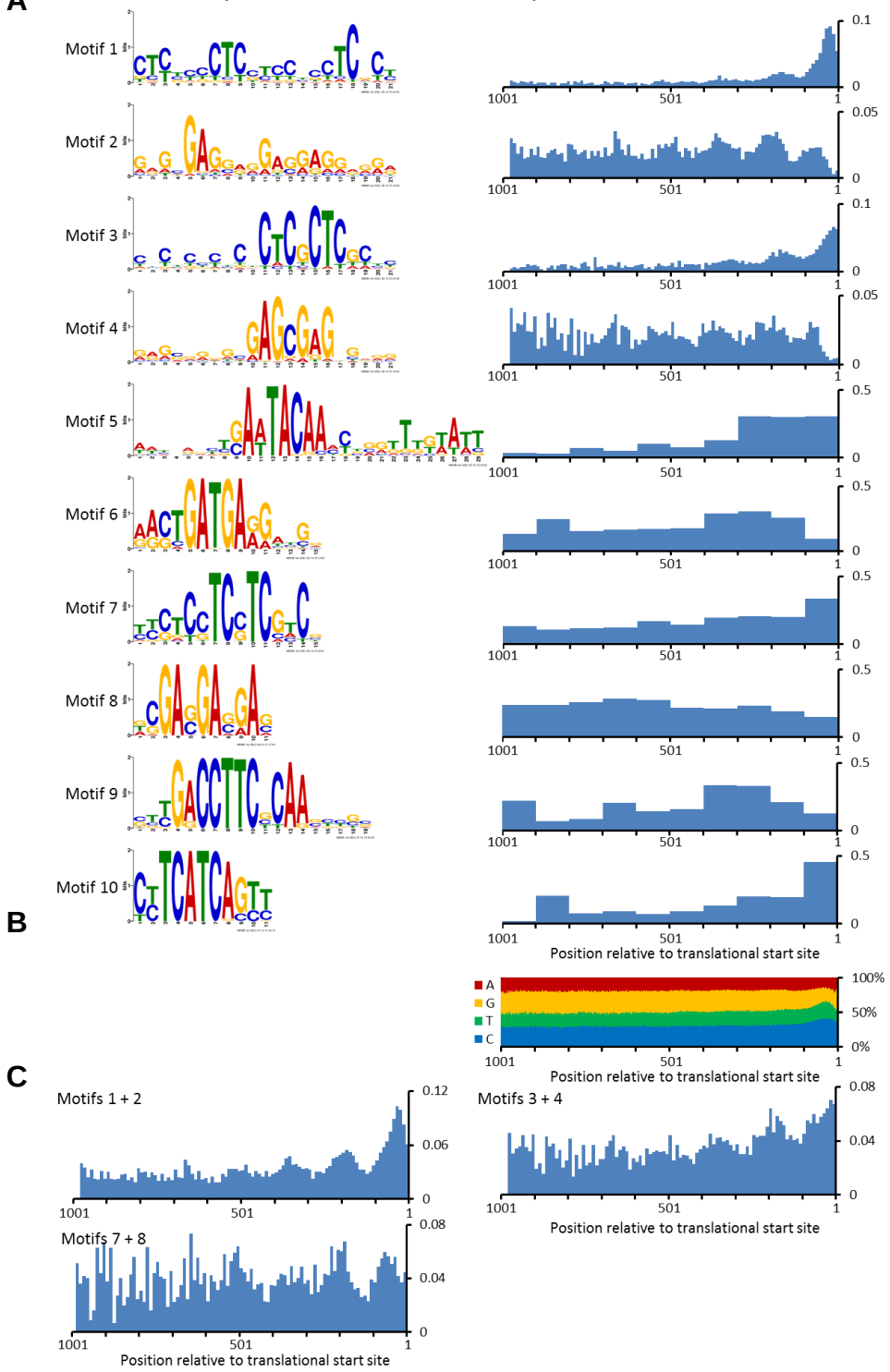


Figure 4.10. Conserved motifs identified in *R. toruloides* promoters. A. 1 kb fragments upstream from all genes in the *R. toruloides* NP 11 genome were subjected to a *de novo* motif search using MEME and the 10 most significant elements reported. LOGO plots on left indicate the motifs found; histograms represent proportion of the total number of each motif found at the positions indicated relative to the translational start site. Motifs 1 and 2 are reverse compliments, as are motifs 3 and 4, 6 and 10 and possibly 7 and 8. **B.** Relative base usage over the same sequence. **C.** proportion of the sum of indicated motifs found at the positions indicated relative to the translational start site.

Excluding their respective reverse compliments the three most common motifs identified in the *R. toruloides* promoter dataset are: the weakly defined CT-rich motif 1, identified in 7215 promoters and most normally found 1-70 bp from the translational start site; the more tightly defined, CT-rich motif 3 identified in 3629 promoters and most commonly identified 1-60 from the transcriptional start site; and the GA- rich motif 8 which was observed in 1053 promoters spread out through the length of promoters but depleted at the translational start site proximal end. In the case of all three motifs their respective complements displayed the opposite pattern, with motifs 2 and 4 appearing throughout promoters but depleted at the translational start site proximal end, and motif 8 enriched at the translational start site proximal end. Motifs 2, 4 and 8 were identified in 5540, 2574 and 950 promoters respectively.

Mean base usage over the length of promoters was calculated and plotted (Figure 4.10B). As expected a bias in favour of C+G relative to A+T was observed over the length of the promoters, but also at between 25 and 45 bp from the translational start site a skew in purine/pyrimidine usage was observed, with around 65 % pyrimidine usage in this region. It was questioned whether this observed skew in purine/pyrimidine usage may be the cause the bias in observed in the relative ratios of the complementary motifs 1 and 2; 3 and 4; and 7 and 8, all of which are purine rich in one direction, and pyrimidine rich in the other. Given the observed pyrimadine usage of 65% shortly upstream of the translational start site, the sequence CTCGCTC, at the core of motif 3, be would expected to be enriched 22 times relative to its reverse complement on the basis of base composition, however in the dataset

motif 3 is only enriched 17 times relative to motif 4, indicating that the pattern observed is likely to be due to the skew in purine/pyrimidine ratio, rather than these motifs causing the skew in base composition.

There appeared to be a second peak in frequencies of motif 1, approximately 180 bp from the transcriptional start site, and three peaks could be observed for motif 3 (Figure 4.10A, right-hand panels). When position data for motifs one and two were summed and plotted this periodic behavior became more evident (Figure 4.10C); these motifs appeared to show a periodicity of around 160 bp. One explanation could be that these motifs are responsible for positioning histones around transcriptional start sites. In order to facilitate access to DNA by RNA polymerase II, histones exist in established positions around the transcriptional start site (Jiang & Pugh, 2009), and the observed periodicity of motifs 1+2 (Figure 4.10C) is consistent with that of nucleosomes in *S. cerevisiae* (Lee et al., 2007) and *S. pombe* (Lantermann *et al.*, 2010). It was noted that these position data are relative to translational start site, rather than the transcriptional start site for which histone location would be important, and 5' UTRs would introduce noise into the data, however the amount of noise introduced would be dependent on the variability in the length of the 5' UTRs. Motifs 3/4, and motifs 7/8 also exhibited periodicity indicating these motifs may also be involved in histone positioning.

Whilst motifs 1/2, 3/4 and 7/8 are may be involved in nucleosome positioning around the transcriptional start site, this does not negate the possible importance of a CT-rich box for translational initiation. Assuming boxes required for translational initiation are degenerate in sequence, they may

overlap with motifs for nucleosome positioning. Therefore, while the observed skew in purine/pyrimidine ratio shortly upstream of the transcriptional start site may be due to a CT-rich region required for translational initiation, the motifs detected by MEME could be achieving a different function (nucleosome positioning) with the same sequence.

Motifs 5, 6, 9 and 10 were not identified in any of the four regulatable promoters studied here and in each case they were only identified in a low number of promoters, (229, 308, 182 and 357) respectively. In each case the top 20 hits were identified and the genes they regulate listed (Table 4.2). In the case of motif 5: five genes are involved in mRNA production, processing and protein production; three in vesicle trafficking and protein maturation, and two are involved in DNA double strand break repair (*MRE11* for NHEJ and *XRCC3* for homologous recombination). One potential explanation could be this element is involved in signalling the transition from S to G2 phase in the cell cycle, up-regulating protein production machinery to facilitate cellular growth and double strand break (DSB) repair machinery to correct any DSBs remaining after DNA replication, prior to cell division (Branzei & Foiani, 2008). Furthermore casein kinase II was identified which is implicated in cell cycle and DNA repair signalling (Litchfield, 2003, Becherel *et al.*, 2010). With regard to motifs 6 and 10, of the 40 genes screened (20 identified for each and then pooled) there were five metabolite transporters and five genes involved with production of ribosomes or tRNA. This could potentially indicate that this motif is involved in a response to growth changes in growth media, (Ljungdahl & Daignan-Fornier, 2012). Finally, in the case of motif 9, no obvious pattern in gene function could be determined to indicate the function of this element.

Table 4.2 Genes downstream of identified promoter motifs 5, 6, 9 and 10.

Motif	Hit	*S. cerevisiae* orthologue	Protein	Notes
5	1		Hypothetical protein	
	2	*MRE11*	Double-strand break repair protein *MRE11*	DNA DSB repair (NHEJ)
	3	*THYN1*	Thymocyte nuclear protein 1	Induced in apoptosis in vertebrates
	4	*TOK1*	Potassium channel subfamily K	Restoration of membrane K^- gradients
	5	*TRX1*	Thioredoxin	Redox signalling
	6	XRCC3	DNA repair protein *XRCC3*	DNA DSB repair (homologous recombination)
	7	*RPS13*	40S ribosomal protein s13	Protein biosynthesis
	8		Hypothetical protein	
	9		Hypothetical protein	
	10	*RRP12*	Ribosomal rRNA processing protein 12	Protein biosynthesis
	11	*FMT1*	Methionyl-tRNA formyltransferase	Mitochondrial protein biosynthesis
	12	*NDE2*	NADH-ubiquinone oxidoreductase 64 kDa subunit	Oxidative phosphorylation
	13	*CKB1*	Caesin kinase II subunit beta	Regulation of cell cycle, DNA repair and others
	14	*VAC8*	Vacuolar protein 8	Sorting of vesicles from Golgi to vacuoles
	15	*RPB11*	RNA polymerase II subunit	Protein biosynthesis
	16	*PRP45*	Pre-mRNA processing protein 45	Protein biosynthesis
	17		Beta-1,4-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase	Protein glycosylation
	18	*TVP38*	SNARE associated Golgi protein	Vesicular trafficking
	19		Zeta toxin p-loop nucleotide triphosphate hydrolase	
	20	*TEA1*	C6 transcription factor	
6	1		Hypothetical protein	
	2	*SRP102*	Signal recognition particle subunit beta	Targeting proteins to the rough endoplasmic reticulum.
	3		Hypothetical protein	
	4	*RPA12*	RNA polymerase 1 subunit	Ribosome biosynthesis
	5	*IN53*	Phosphatidylinositol phosphate phosphatase	Intracellular signalling
	6		T-complex 11 family protein	Intracellular signalling
	7	*DBP4*	ATP-dependent RNA helicase	Ribosome biogenesis
	8		Hypothetical protein	
	9	*GCN5*	Histone acetyltransferase	Gene regulation
	10	*TRM112*	tRNA methyltransferase m2G10	tRNA maturation
	11	*SFA1*	Alcohol dehydrogenase	
	12	*SFA1*	Alcohol dehydrogenase	
	13		Major facilitator superfamily (MFS) protein	Metabolite uptake
	14	*TNA1*	MFS nicotinic acid transporter	Metabolite uptake
	15		MFS transporter	Metabolite uptake
	16		C6 family transcription factor	Gene regulation
	17		N-acetyltransferase	
	18		Hypothetical protein	
	19	*SEN54*	tRNA splicing endonuclease subunit	tRNA maturation
	20		MFS transporter	Metabolite uptake

9	1		Hypothetical protein	
	2		Hypothetical protein	
	3		Hypothetical protein	
	4		Hypothetical protein	
	5		Hypothetical protein	
	6		Hypothetical protein	
	7	*SLN1*	Sensor histidine kinase VicK	Two-component signalling system
	8		Hypothetical protein	
	9	*ENV9*	short-chain dehydrogenase/reductase SDR family protein	
	10		acyl-CoA N-acyltransferase domain containing protein	
	11		Hypothetical protein	
	12		Expansin family protein	Cell growth
	13		Hypothetical protein	
	14		Proteophosphoglycan ppg4	
	15		Hypothetical protein	
	16		Hypothetical protein	
	17		Hypothetical protein	
	18		Methylmalonate-semialdehyde dehydrogenase	
	19		Hypothetical protein	
	20	*ZRT1*	ZIP-like iron-zinc transporter	Metal uptake
10	1		Hypothetical protein	
	2		Hypothetical protein	
	3		Transcriptional regulator	Gene regulation
	4	*LFA38*	Beta-keto reductase	
	5		Hypothetical protein	
	6		Zinc finger, MYND-type domain containing protein	
	7	*GCN20*	Iron complex transport system ATP-binding protein	Metal homeostasis
	8		Hypothetical protein	
	9		zinc finger, MYND-type domain containing protein	
	10		zinc finger, MYND-type domain containing protein	
	11		Acetate kinase	
	12		Zinc finger, RING-CH-type protein	
	13		RAI1-like domain containing protein	rRNA maturation (possibly)
	14		Modifier of rudimentary, Modr domain protein	Gene regulation
	15		Nucleus protein	
	16	*RGA1*	Rho GTPase activating protein	Intracellular signalling / cytoskeleton
	17		Hypothetical protein	
	18	*PUT4*	Amino acid transmembrane transporter	Metabolite uptake
	19		Hypothetical protein	
	20		Hypothetical protein	

This *de novo* motif search was far from exhaustive and does not rule out the possibility of conserved eukaryotic promoter motifs such as TATA or CAAT boxes playing a role in gene regulation in *R. toruloides*. However, were such elements highly conserved and found in a majority of promoters we would expect to have identified these elements in this *de novo* motif search before more weakly conserved, or elements only represented in a low number of promoters such as some of the elements identified.

4.3 Conclusion

In order to measure promoter activity, expression of EGFP and Venus YFP marker genes were demonstrated in *R. toruloides* CBS 14. Using these and the tools developed in chapter 3, four inducible promoters have been characterised to allow controllable expression in the oleaginous yeast *R. toruloides* CBS 14. The *NAR1* promoter is strongest when induced, shows tight regulation under repressed conditions in two *R. toruloides* strain backgrounds, has a short 200 bp functional sequence, and would be the first choice promoter in many cases. However, each promoter has its own individual characteristics that render it suitable for particular applications, and together they provide a suite of complementary regulatory elements for controlling gene expression in this yeast. No common conserved core promoter elements were identified after performing a *de novo* motif search of all promoters from *R. toruloides*, however, the periodic occurrence of degenerate, CT rich motifs potentially indicates that histone positioning may be important for gene regulation in *R. toruloides*.

The development of tools for molecular genetic analysis of *R. toruloides* CBS 14 should permit metabolic engineering or this oleaginous yeast. The next chapter describes prof of principle experiments seeking production of hydrocarbons in this strain.

5 *R. toruloides* CBS 14 as a platform for hydrocarbon biosynthesis

5.1 Introduction

As discussed in section 1.1.2, under carbon replete, nitrogen limited conditions, *R. toruloides* accumulates lipids up to 76 % of its dry biomass, the majority as triacylglycerols (TAGs) (Li *et al.*, 2006, Fei *et al.*, 2016). Moreover, through over-expression of the native acetyl-CoA carboxylase and diacylglycerol acyltransferase genes, lipid production can be further increased (Zhang *et al.*, 2016). These lipids and derivative compounds are potentially valuable commodities, and could be used for production of sustainable, 'drop-in' petroleum replacement fuels (Steen *et al.*, 2010, Howard *et al.*, 2013, da Silva *et al.*, 2014, Liao *et al.*, 2016).

Biodiesel is currently produced by transesterification of different lipids, mostly from oleaginous plants, most commonly using methanol or ethanol to yield fatty acid methyl esters (FAMEs) or fatty acid ethyl esters (FAEEs) respectively. However, the high cost of the feedstock, and concerns over food vs. fuel limits their uptake (Haas *et al.*, 2006, Lee *et al.*, 2015). Microbial production can use more cost effective, land use-efficient feedstocks (da Silva *et al.*, 2014, Muniraj *et al.*, 2015), and expression of wax ester synthase means fatty esters can been produced *in vivo* (Kalscheuer *et al.*, 2004, Shi *et al.*, 2014) (Figure 5.1), reducing the cost of downstream processing. However *in vivo* transesterification does not overcome the adverse physical properties of biodiesel: fatty esters are hygroscopic, susceptible to low temperature

gelling, oxidatively unstable and can attack certain materials used in the fuel systems of most vehicles, limiting their usefulness (Fazal *et al.*, 2011). As a result, for commercial use biodiesel must be blended with petrodiesel before sale; for example in Europe the maximum blend sold at retail filing stations is B7, which contains 7 % fatty esters (Kampman *et al.*, 2013).

Petrodiesel contains a mixture of hydrocarbons, including approximately 75 % alkanes (including linear, branched and cycloalkanes) and 25 % aromatic hydrocarbons, with molecules containing between 8 and 21 carbon atoms. Due to the large amount of pre-existing infrastructure, the best straightforward replacement for this would be to produce a mixture identical in composition, but produced in a renewable, carbon-neutral manner. Whilst it would be difficult to biologically replicate the aromatic fraction of petroleum derived fuels, exclusion of aromatic compounds from fuels can reduce emissions of unburnt hydrocarbons (Barbella *et al.*, 1989), and through appropriate blending of biologically sourced alkanes or alkenes a carbon-neutral, drop-in fuel could be produced with properties equivalent or superior to contemporary petrodiesel (Howard *et al.*, 2013, Liao *et al.*, 2016).

5.1.1 Pathways for *in vivo* production of hydrocarbons

Native hydrocarbon production has been reported in plants, insects, bacteria and cyanobacteria (Herman & Zhang, 2016). Many of these pathways and variations thereof have been recapitulated *in vitro* or *in vivo*, and have potential for production of ‘drop-in’ fuels (Schirmer *et al.*, 2010, Rude *et al.*, 2011, Howard *et al.*, 2013, Liu *et al.*, 2014b, Gupta & Phulara, 2015).

Before elucidation of pathways for lipid derived hydrocarbons, isoprenoids garnered interest as potential replacements for petroleum-derived fuels (Fortman *et al.*, 2008), however low yields have limited their development (Gupta & Phulara, 2015). In 2009 Beller *et al.* described a mechanism by which long chain hydrocarbons can be produced by head-to-head condensation of fatty acyl-CoAs, and were able to use this system to produce long chain medial alkenes in *E. coli*; however such long chain hydrocarbons would not be well suited to production of fuels. Since publication of this work, other metabolic pathways have been reported which are better suited to synthesis of molecules for use as a 'drop-in' replacement for petrodiesel without a requirement for downstream cracking.

In 2010 Schirmer *et al.* described the pathway by which cyanobacteria produce alkanes from fatty acids, through reduction of fatty acyl-ACP to a fatty aldehyde by the action of the enzyme acyl-ACP reductase, before decarbonylation of the resulting fatty aldehyde by the action of the enzyme aldehyde decarbonylase liberating an alkane with one less carbon than the starting fatty acid. Furthermore, they demonstrated that heterologous expression of the acyl-ACP reductase and aldehyde decarbonylase gave alkane biosynthesis in *E. coli*. Following the publication of this work, variations and improvements on this pathway have been published, including generation of fatty aldehydes from fatty acyl-CoA (Steen *et al.*, 2010) or from free fatty acids (FFAs) (Akhtar *et al.*, 2013, Howard *et al.*, 2013), increasing concentrations of starting materials (Liu *et al.*, 2014a, Zhou *et al.*, 2016), or transplanting the pathways developed into organisms better suited to

production of biofuels, including into the oleaginous yeast *Yarrowia lipolytica* (Xu *et al.*, 2016).

While most developments in biosynthesis of hydrocarbons for biofuels have been improvements or variations on the pathway of Schirmer *et al.* (2010), research on alternative pathways for hydrocarbon biosynthesis has continued. Rude *et al.* (2011) identified a cytochrome p450 enzyme responsible for the hydrogen peroxide-dependent one-step decarboxylation of fatty acids, liberating terminal alkenes of length one less than the starting fatty acid in *Jeotgalicoccus sp.* This enzyme, named OleT, has the advantage that heterologous expression of only a single gene is required for engineered production of hydrocarbons (Rude *et al.*, 2011).

Furthermore OleT uses H_2O_2 as a source of reducing potential. Whilst this may be problematic for scale up production, this is advantageous in proof of principle studies. Production of alkanes from fatty aldehydes by *Synechococcus elongatus* aldehyde decarbonylase uses ferredoxin as a source of reducing potential; in eukaryotes this is problematic as ferredoxin is compartmentalised into mitochondria, and therefore, for functional expression of aldehyde decarbonylase in *S. cerevisiae*, co-expression of cytoplasmic bacterial ferredoxin and ferredoxin reductase is required (Chen *et al.*, 2015). Whilst H_2O_2 is primarily produced in mitochondria as a side product of the electron transport chain, it can relatively freely diffuse across membranes and is therefore available to cytoplasmic OleT.

FFAs can be cytotoxic (Eisenberg & Buttner, 2014) and are maintained at low levels within cells, with excess lipid stored as TAGs; reported alkene yields

are low (Rude *et al.*, 2011), in part due to the lack of substrate (Liu *et al.*, 2014b, Chen *et al.*, 2015). Therefore, to increase alkene yield, OleT has been expressed in cells engineered to increase the availability of FFAs in both *S. cerevisiae* and *E. coli* (Liu *et al.*, 2014b, Chen *et al.*, 2015). While an inability to perform targeted integration limits our ability to perform metabolic engineering of *R. toruloides*, genes to increase the concentration of FFAs can be expressed. In *E. coli*, OleT was expressed in conjunction with overexpression of the native acyl-CoA thioesterase I (TesA) (Liu *et al.*, 2014a); this has the dual purpose that it alleviates product inhibition of FAS, upregulating lipid production, and increases the proportion of lipid as FFA. Unlike in *E. coli*, in *R. toruloides* the rate of fatty acid synthesis is unlikely to be a limiting factor in production of alkenes, but the concentration of FFAs may be. Once lipids are partitioned as TAGs they are inaccessible to OleT or TesA. *Thermomyces lanuginosus* lipase II (Lip2) catalyses the hydrolysis of TAGs to FFAs and glycerol (Fernandez-Lafuente, 2010). This activity provides a greater pool of substrate for production of alkenes. Co-expression of *OleT* and Lip2 has been used to increase the yield of alkenes from exogenously supplied TAGs *in vitro* and in a whole cell system (Yan *et al.*, 2015), however this has not been demonstrated in a system concurrent with lipid production.

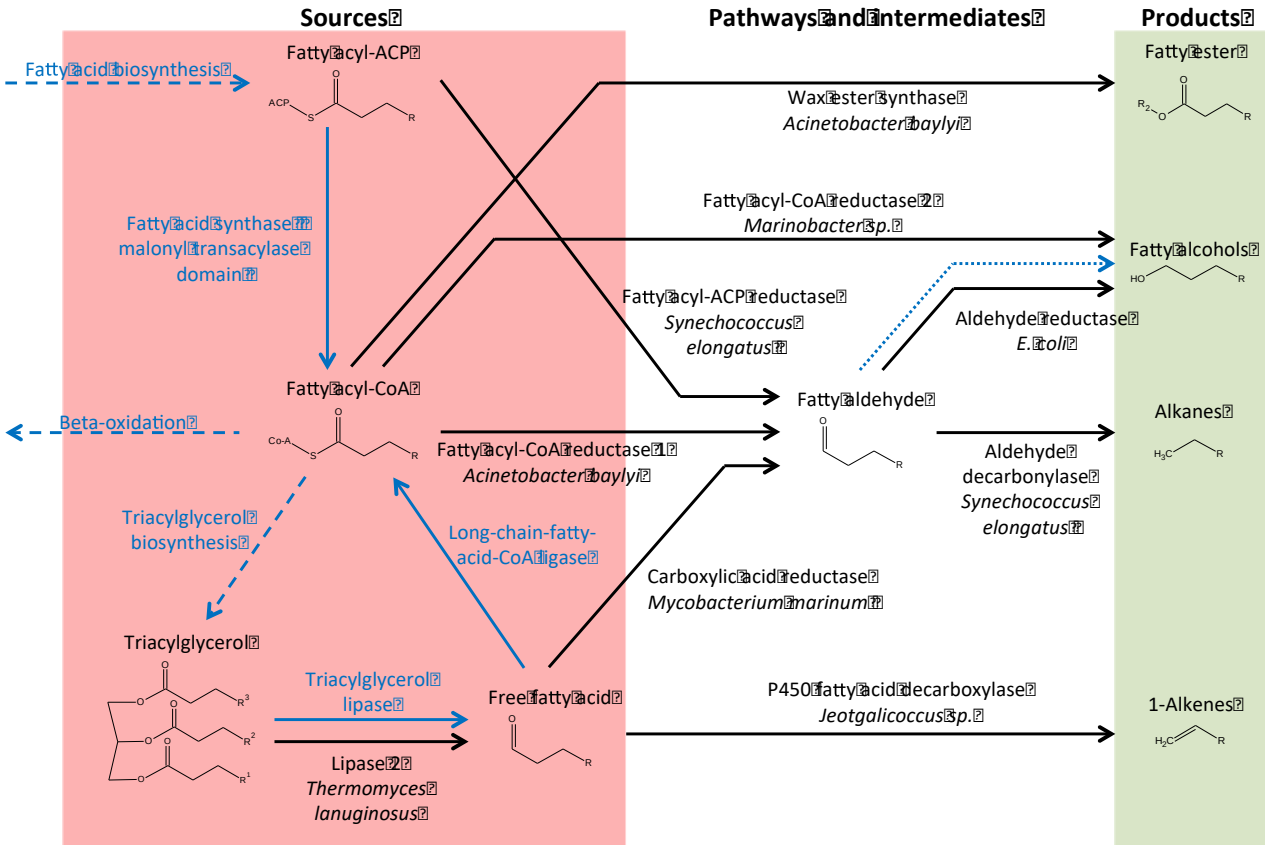


Figure 5.1 Pathways for *in vivo* production of chemicals from fatty acids. Source compounds are highlighted in red and product compounds in green. Pathways preexisting in *R. toruloides* are shown in blue; pathways used for production of biofuels or other molecules from fatty acids are in black. Dashed lines indicate multi-step pathways. An example source organism is indicated for heterologously expressed enzymes. Native reduction of fatty aldehydes to fatty alcohols is shown by a dotted line as this can be the result of promiscuous activity of many enzymes (Zhou *et al.*, 2016).

5.1.2 Section aims

For the reasons discussed in section 1.1, *R. toruloides* would make a good host organism for production of biodiesel or other lipid-derived products. Therefore the aim of this chapter was to demonstrate production of hydrocarbons in this organism as a proof of principle for production of 'drop-in' petroleum replacement diesel fuel by *R. toruloides*. In order to achieve this end the production of alkenes would be sought by expression of OleT in *R. toruloides*, co-expressed with Lip2. Furthermore it was hoped to demonstrate production of alkanes by expression of two variations on the pathway of Schirmer *et al.* (2010): (a) with co-expression of fatty acyl-CoA reductase and aldehyde decarbonylase in order to produce alkanes from fatty acyl-CoA; and (b) co-expression of fatty acyl-ACP reductase and aldehyde decarbonylase in order to produce alkanes from fatty acyl-ACP.

5.2 Results and discussion

5.2.1 Expression of OleT in *R. toruloides*

Because of its simplicity, only requiring expression of a single gene, hydrocarbon biosynthesis in *R. toruloides* was first attempted by expression of OleT. To maximise alkene biosynthesis in *S. cerevisiae*, Chen *et al.* (2015) screened seven orthologues of the OleT gene from different bacteria and determined that a codon-optimised *Jeotgalicoccus sp.* ATCC 8456 OleT achieved the highest activity under the conditions tested. An *R. toruloides* codon optimised *Jeotgalicoccus sp.* ATCC 8456 OleT gene was therefore synthesised and inserted into *Eco*RI/*Afl*II-digested pEGFP-Rt-YR-G418 by in-

yeast assembly to give plasmid pOleT-Rt-YR-G418 in which the EGFP gene of pEGFP-Rt-YR (Figure 3.8) has been replaced by *OleT.* This construct was transformed into *R. toruloides* CBS 14.

As hydrocarbons are produced from lipid (Figure 5.1), in order to maximise potential hydrocarbon yield, cultures were grown under lipid accumulating conditions. Lipid accumulation can be stimulated by growth in carbon replete, nitrogen limited conditions, however nitrogen is still required for growth therefore cannot be removed completely. Wiebe *et al.* (2012) measured lipid accumulation by *R. toruloides* CBS 14 with different concentrations of sugar, different carbon/nitrogen (C/N) ratios and by culturing cells for different lengths of time; based on these observations, cultures were grown in shake flasks at 30 °C for three days, in YNB media modified to contain a final glucose concentration of 30 g L^{-1} with a carbon/nitrogen ratio of 65. This C/N ratio was selected as decreasing the C/N results in a decrease in lipid accumulation, however further increasing the C/N ratio results little further change in total lipid accumulation at, the cost of biomass and the rate at which maximum lipid content is achieved (Wiebe *et al.*, 2012). When I grew wild type *R. toruloides* CBS 14 in this medium for three days a final lipid content of around 50 % dry cell weight was observed (Figure 5.4) which is in line with the observations of Wiebe *et al.* (2012).

50 mL *R. toruloides* cultures expressing OleT, as well as wild type controls were grown for three days under the lipid accumulating conditions previously described. Cells were harvested and re-suspended in 1 mL methanol with 0.1 % tetrabutylammonium hydroxide (TBAH) and 25 ng μL^{-1} 1-tetradecene as an

internal standard. Cells were lysed by bead-beating and hydrocarbons extracted by solvent extraction into hexane. Samples were analysed by gas chromatography-mass spectrometry (GC-MS) in order to detect hydrocarbons, however in no cases were hydrocarbons other than the 1-tetradecene internal standard observed.

Re-inspection of the sequence of the codon-optimised OleT gene showed the presence of two potential 5' splice sites. To avoid any possibility of incorrect splicing these sites were therefore removed by site-directed point mutation, a hexahistidine tag was engineered at the N-terminus of the protein, and expression was verified by western blot (Figure 5.2). The western blot indicated OleT was expressed both with and without the putative 5' splice sites removed, but that removal of the 5' splice sites increased expression of the OleT gene.

In order to try and detect alkenes produced by the action of OleT expressed from coding sequence with putative splice sites removed, growth under lipid accumulating conditions was repeated and extracted samples were analysed by GC-MS (Figure 5.3). The experiment was performed in biological triplicate with independent transformants and for one transformant chromatograph peaks were observed at retention times of 569, 657.5 and 667 s on extracted ion chromatograms for a 55.05 mass fragment (Figure 5.3). The peak at 667 s corresponds to 1-heptadecane, produced by the decarboxylation of stearic acid (C18:0) and the peak at 657.5 s potentially corresponds to 1,8-heptadecadiene produced by the decarboxylation of oleic acid (C18:1). Standards to confirm the GC retention time of 1,8-heptadecadiene were

unavailable and therefore the identity of this 657.5s peak cannot be confirmed. Furthermore, in two of the transformant lines a small peak was observed at 569 s corresponding to 1-pentadecene, which would be produced by the decarboxylation of palmitic acid (C16:0).

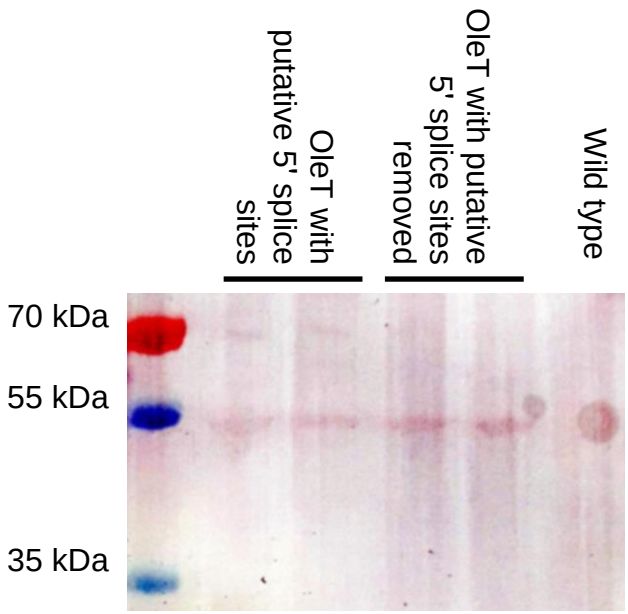


Figure 5.2. Western blot to confirm expression of OleT in *R. toruloides*. A hexahistidine tag sequence was engineered on to the N terminus of the codon optimised OleT open reading frame and point mutation was performed to remove putative 5' splice sites identified in the synthetic gene. Protein extraction and Western blotting using hexahistidine specific antibodies was performed on *R. toruloides* transformed with his-tagged OleT with or without putative 5' splice sites removed. In lines transformed with either OleT variant, a band was observed at approximately 55 kDa (expected molecular mass 50 kDa), but was not observed in untransformed *R. toruloides*, indicating expression of OleT. For both OleT variants protein extraction was performed in duplicate with independently transformed strains.

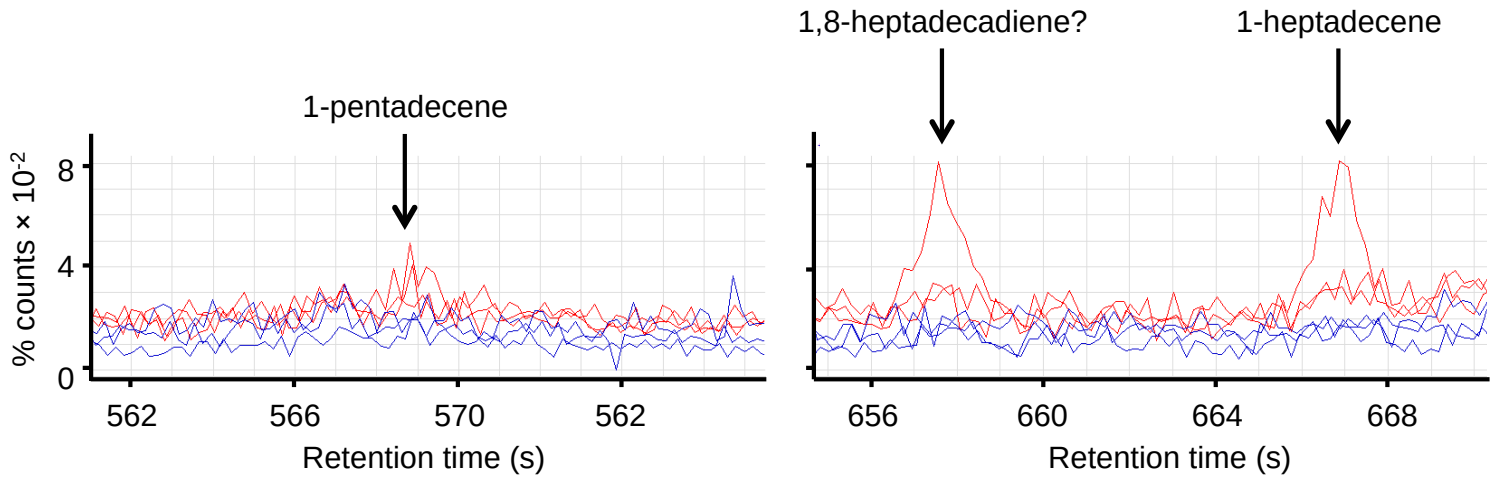


Figure 5.3. Production of alkenes in *R. toruloides* by expression of OleT. *R. toruloides* expressing codon optimised OleT with putative 5' splice sites removed (red) and wild type *R. toruloides* (blue) were grown in lipid accumulating conditions for three days before hydrocarbons were extracted by solvent extraction into hexane. The organic fraction was analysed by GC-MS and extracted ion chromatograms for a 55.05 mass fragment plotted for each sample. Experiments were performed in biological triplicate, with independent transformants.

In this experiment apparent alkene biosynthesis was not consistent between the three transformant strains, and where alkenes were observed they were at very low concentrations. Confirming whether the alkenes observed were the result of biological hydrocarbon biosynthesis would require further work including stable-isotope tracer experiments.

5.2.2 Expression of *Thermomyces lanuginosus* lipase II

In order to increase the concentration of FFAs within cells and therefore increase available substrate for OleT, the lipase *Lip2* was co-expressed with OleT. An *R. toruloides* codon optimised Lip2 was synthesised and cloned into *Eco*RI/*Afl*II-digested pEGFP-Rt-YR-Hyg (Figure 3.8) by in-yeast assembly to give plasmid pLip2-Rt-YR-Hyg. By placing the construct harbouring OleT under G418 selection and the Lip2 construct under hygromycin selection this allows single transformation of yeast with either pOleT-Rt-YR-G418, pLip2-Rt-YR-Hyg, or sequential transformation with both plasmids.

To determine whether expression of Lip2 increased the concentration of FFAs in *R. toruloides* cells, the method of Brown (2016) was used to determine the concentration of FFAs. Briefly, 100 µL culture sample was taken and 1 µL of a 5 g L^{-1} heptadecanoic acid standard added. Cells were lysed by bead-beating, followed by treatment with TBAH. FFAs were methylated by treatment with iodomethane as this gives high specificity for methylation of FFAs relative to other lipid species yielding FAMES (Patterson *et al.*, 1999), and the resulting FAME concentrations were measured by gas chromatography-flame ionising detection (GC-FID). Methyl esters of palmitic (C16:0), stearic (C18:0), oleic

(C18:1) and linoleic acids (C18:2) were measured (Figure 5.4A). This showed a significant increase in the concentration of all FFAs measured (t-test, $p > 0.05$) and a 1.3-fold increase in total FFA content. It was noted that Lip2 did not increase the concentration of FFAs equally, and that the relative increase in the concentrations was stearic acid > palmitic acid > oleic acid >> linoleic acid. The reported substrate preference of Lip2 can go some way to explaining this differential increase in the different fatty acids as Lip2 has a lower reported activity against poly-unsaturated fatty acids relative to saturated or monounsaturated fatty acids (Moharana *et al.*, 2016).

FFAs are potentially cytotoxic, act as signalling molecules within cells and partitioning of lipid can influence lipid biosynthesis. Therefore dry cell mass and total lipid content were measured gravimetrically, and % lipid content calculated for wild type cells, and cells expressing Lip2 after three days growth in lipid accumulating conditions (Figure 5.4B-D). No significant differneces were observed for dry cell weight, total lipid content or lipid fraction (T-test $p > 0.05$), indicating no major cytotoxic or lipid biosynthesis effects of Lip2 expression in *R. toruloides.*

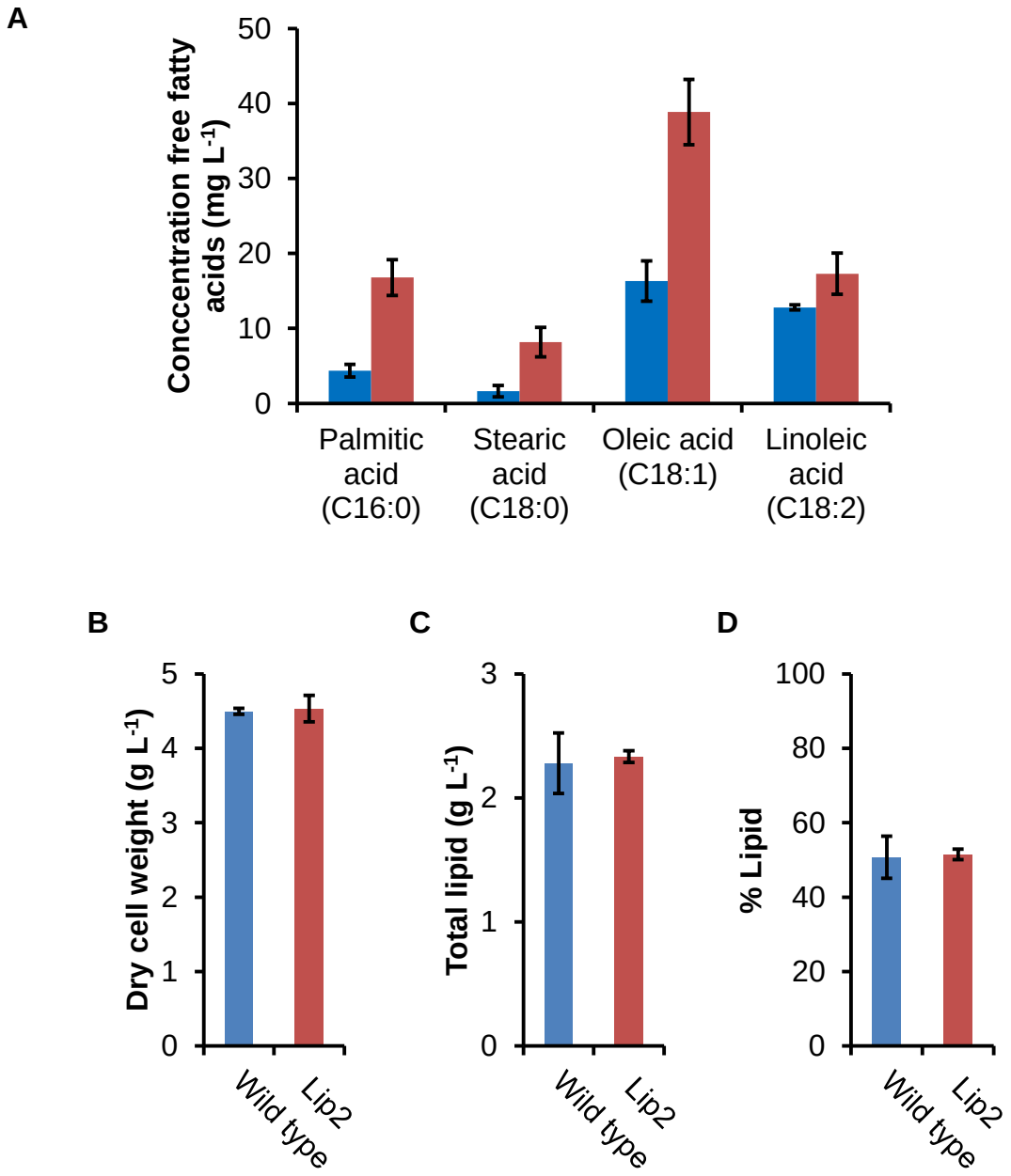


Figure 5.4. Effect of expression of *T. lanuginosus* lipase II (Lip2) on free fatty acids and total lipids in *R. toruloides* CBS 14. A. *R. toruloides* cultures with and without Lip2 were grown for three days under lipid accumulating conditions before FFA concentrations were measured by GC-FID. Bars show the mean of three biological replicates (independent transformants); error bars indicate standard deviation **B-D.** Dry cell weight, total lipid determined gravimetrically and % lipid content of *R. toruloides* cultures with and without Lip2 after three days' growth under lipid accumulating conditions. Bars show the mean of three biological replicates (independent transformants); error bars indicate standard deviation.

Lip2 was co-expressed with OleT in *R. toruloides*, and alkene biosynthesis measured as before (Figure 5.5). In all three double transformant cell lines expressing both OleT and Lip2, small peaks were observed at elution times of 667 s and 657.5 s corresponding to 1-heptadecene and (potentially) 1,8-heptadecadiene. No other peaks were observed relative to strains singly transformed with Lip2. These data indicate the production of alkenes in *R. toruloides* expressing OleT, however the hydrocarbons observed were at very low concentrations and the data provide no evidence that co-expression of Lip2 affects alkene biosynthesis.

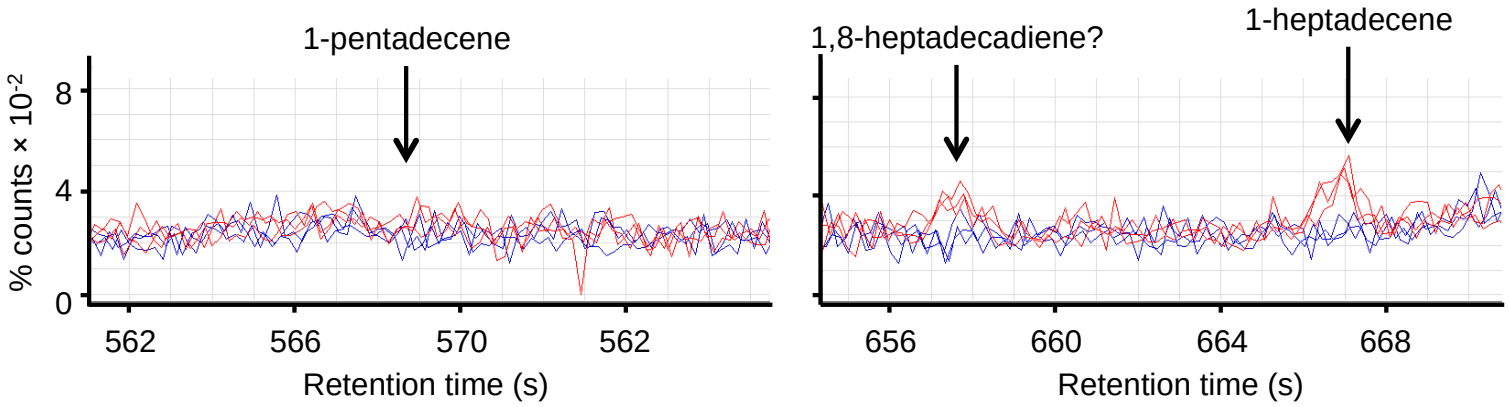


Figure 5.5. Production of alkenes in *R. toruloides* by co-expression of OleT and Lip2. *R. toruloides* expressing codon optimised OleT and Lip2 (red) and *R. toruloides* only expressing Lip2 (blue) were grown in lipid accumulating conditions for three days before hydrocarbons were extracted by solvent extraction into hexane. The organic fraction was analysed by GC-MS and extracted ion chromatograms for a 55.05 mass fragment plotted for each sample. Experiments were performed in biological triplicate, with independent transformants.

5.2.3 Expression of alkane biosynthesis pathways in *R. toruloides*

Alkanes can be produced *in vivo* by decarbonylation of fatty aldehydes (Schirmer *et al.*, 2010) which, in turn are produced by reduction of fatty acids or their conjugates. Within cells there are three lipid pools which can be exploited to produce fatty aldehydes: fatty acyl-ACPs, which can be reduced by the action of the enzyme fatty acyl-ACP reductase (Schirmer *et al.*, 2010); fatty acyl-CoAs which can be reduced by the action of the enzyme fatty acyl-CoA reductase I (Steen *et al.*, 2010); or FFAs which can be reduced by the action of the enzyme carboxylic acid reductase (Akhtar *et al.*, 2013, Howard *et al.*, 2013) (Figure 5.1). These three enzymes are monomeric and use NADPH as a source of reducing potential (carboxylic acid reductase also requires ATP). As NADPH is not compartmentalised to any particular subcellular location, unlike fatty aldehyde decarbonylase neither fatty acyl-ACP reductase, fatty acyl-CoA reductase or carboxylic acid reductase require expression of a supplementary reduction system for cytoplasmic activity in eukaryotes. This property means that fatty acyl-ACP reductase and fatty acyl-CoA reductase only require expression of a single gene for *in vivo* production of fatty aldehydes; carboxylic acid reductase however also requires expression of a phosphopantetheinyl transferase in order to catalase addition of a phosphopantetheine cofactor required for activity (Akhtar *et al.*, 2013).

Decarbonylation of fatty aldehydes to alkanes is typically catalysed by cyanobacterial fatty aldehyde decarbonylase which requires reduced ferredoxin as a co-factor (Schirmer *et al.*, 2010). In eukaryotes ferredoxin is sequestered into mitochondria. The *R. toruloides* genome contains two

ferredoxin genes (Zhu *et al.*, 2012) and protein sequence analysis confirmed that mitochondrial targeting sequences are present on both *R. toruloides* ferredoxin proteins (Claros & Vincens, 1996). For activity of aldehyde decarbonylase in *S. cerevisiae* co-expression cytoplasmic with *E. coli* ferredoxin and ferredoxin reductase was required (Buijs *et al.*, 2015). Therefore, it was intended to co-express these three proteins together in *R. toruloides*.

Synthetic constructs were designed such that a codon optimised gene cluster encoding *S. elongatus* aldehyde decarbonylase, *E. coli* ferredoxin and ferredoxin reductase under the regulation of the *R. toruloides* CBS 14 *TUB1* (alpha tubulin), *THI5* and *THI4* constitutive promoters respectively could be transformed into *R. toruloides* with or without *S. elongatus* fatty acyl-ACP reductase or *Acinetobacter baylyi* fatty acyl-CoA reductase sequences under the regulation of the *R. toruloides* CBS 14 *PGK1* promoter (Figure 5.6).

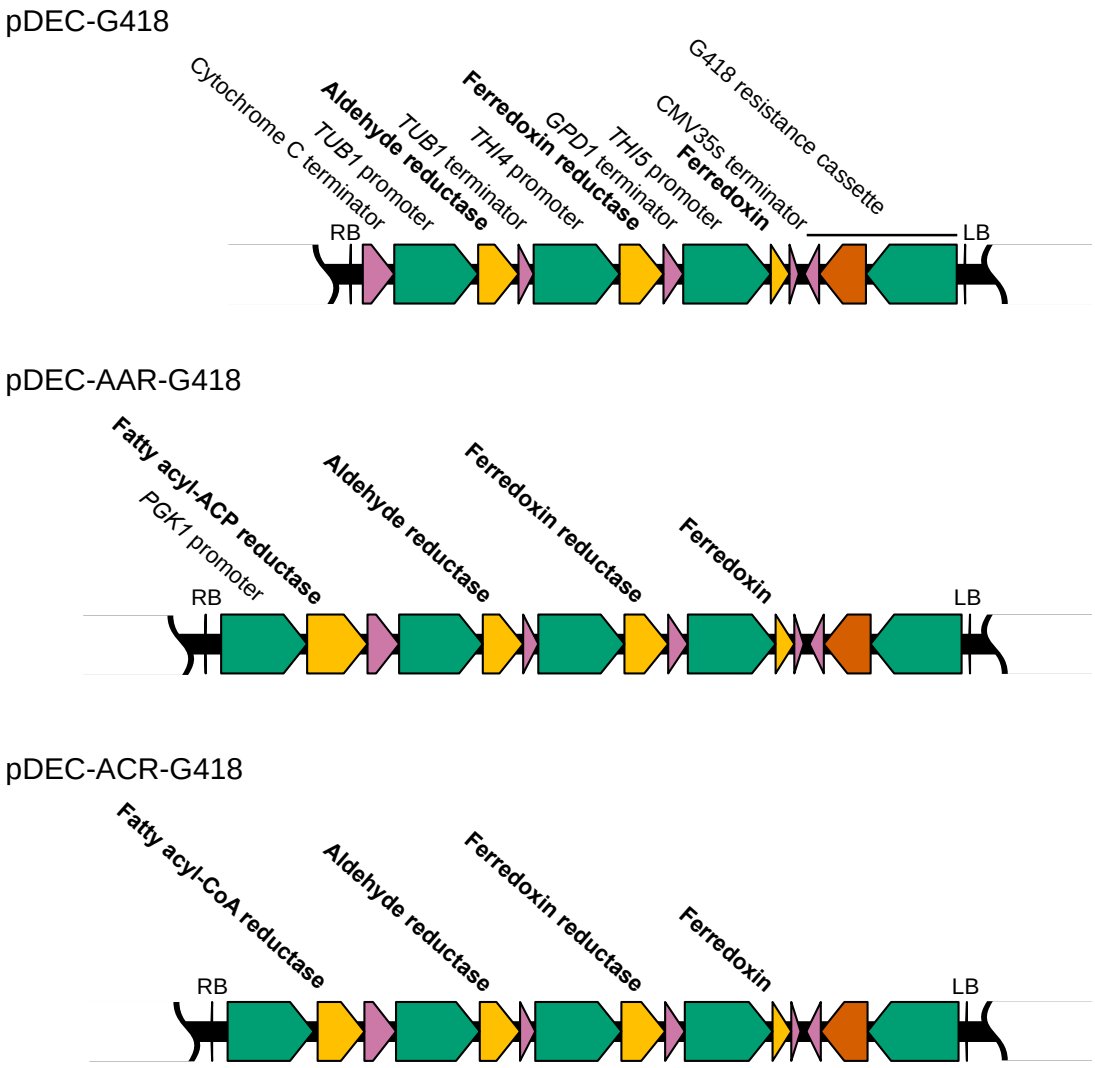


Figure 5.6. T-DNA region of constructs for alkane biosynthesis in *R. toruloides*. A synthetic construct containing codon optimised aldehyde decarbonylase, ferredoxin and ferredoxin reductase sequences under regulation of the *R. toruloides TUB1*, *THI5* and *THI4* constitutive promoters respectively was inserted into *Pml*I/*Spe*I-digested pEGFP-Rt-YR-G418 by in-yeast assembly to give pDEC-G418. Constructs pDEC-AAR-G418 and pDEC-ACR-G418 were similarly constructed and additionally contained codon optimised sequences encoding fatty acyl-ACP reductase (AAR) or fatty acyl-CoA reductase (ACR) respectively, under regulation of the *PGK1* constitutive promoter. Green segments indicate promoter sequences, yellow indicates coding sequence of genes for alkane biosynthesis, purple indicates terminators and orange indicates antibiotic resistance marker.

50 mL cultures of wild type *R. toruloides* or *R. toruloides* transformed with plasmids pDEC-G418, pDEC-AAR-G418 or pDEC-ACR-G418 were grown under the lipid accumulating conditions described in section 5.2.1 for three days. Cells were harvested, re-suspended in 1 mL methanol with 0.1 % tetrabutylammonium hydroxide (TBAH) and 25 ng μL^{-1} tetradecane as an internal standard, lysed by bead-beating, and hydrocarbons extracted by solvent extraction into hexane. Hydrocarbons were then measured by GC-MS.

In cultures of wild type *R. toruloides* or cells transformed with pDEC-G418 or pDEC-ACR-G418 no evidence for hydrocarbons could be detected other than three small peaks observed with a retention time of around 541 s, the centre of which corresponds to the retention time of hexadecane. These peaks were observed in all samples and the mass spectra indicated these were unlikely to be alkanes, but rather a co-eluting contaminant. However in cultures transformed with plasmid pDEC-AAR-G418, a peak at 628 s corresponding to heptadecane could be detected (Figure 5.7A). The identity of this peak was confirmed by its mass spectrum (Figure 5.7B). This heptadecane would result from reduction and subsequent de-carbonylation of stearic acid conjugated to the ACP domain of *R. toruloides* FAS. This is consistent with the previously-reported expression of fatty acyl-ACP reductase and fatty aldehyde decarbonylase in *S. cerevisiae*, which yielded 2.7±0.9 μg DW^{-1} heptadecane but no other observable alkanes (Buijs *et al.*, 2015).

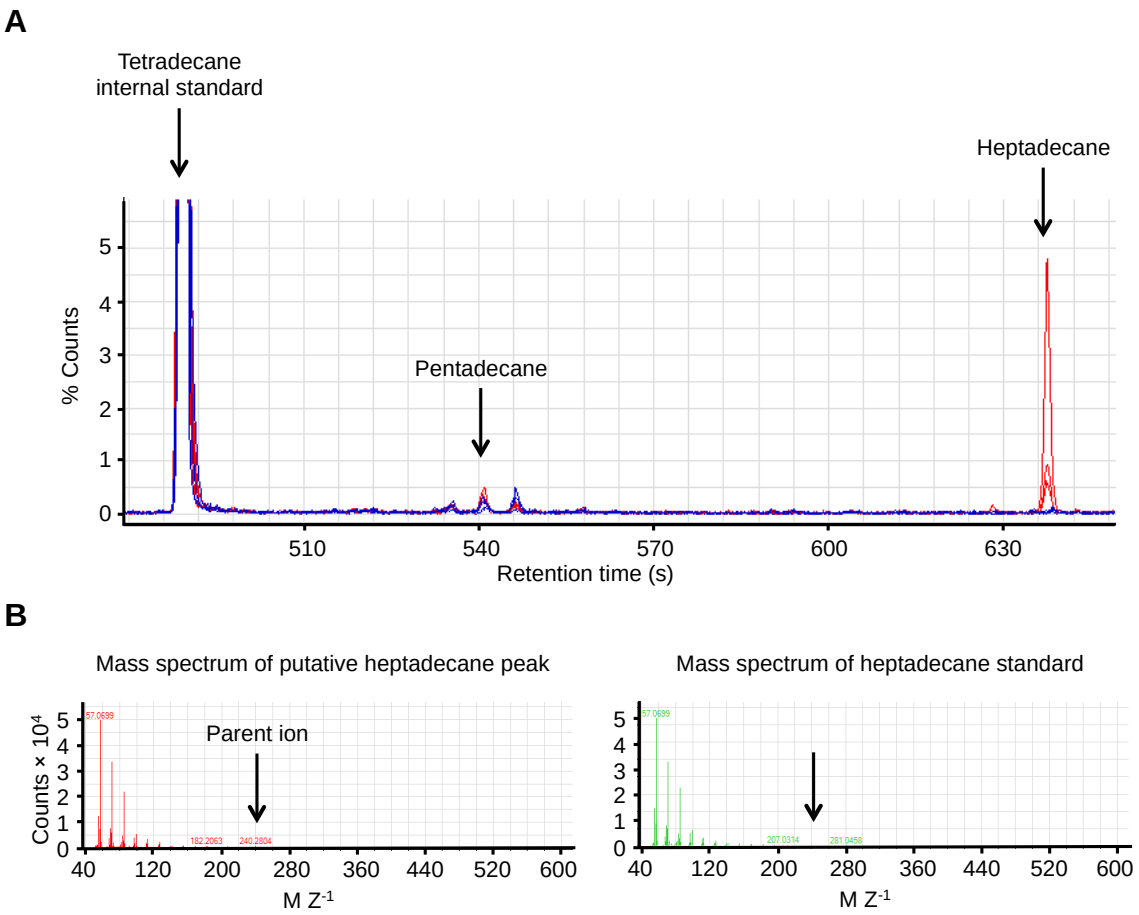


Figure 5.7. Production of heptadecane in *R. toruloides* CBS 14. A. *R. toruloides* transformed with pDEC-AAR-G418 (red) and wild type *R. toruloides* (blue) were grown under lipid accumulating conditions for three days and hydrocarbons produced were analysed by GC-MS. **B.** Comparison of mass spectra for the observed heptadecane peak and a known heptadecane standard to confirm the identity of the peak.

5.3 Conclusion

These investigations found production of heptadecane in *R. toruloides* CBS 14 through co-expression of *S. elongatus* fatty acyl-ACP reductase and fatty aldehyde decarbonylase, and *E. coli* ferredoxin and ferredoxin reductase. Furthermore trace production of alkenes was observed in *R. toruloides* CBS 14 by expression of the *Jeotgalicoccus sp.* fatty acid decarboxylase OleT. To confirm that the observed hydrocarbons were of biological origin would require further work, for example feeding with isotope labelled glucose. Additionally, an increase in the concentration of FFAs within *R. toruloides* CBS 14 after expression of *T. lanuginosus* lipase II (Lip2) was observed. Unfortunately no increase in the alkene yield from OleT by co-expression with Lip2 could be demonstrated, but this enzyme may still prove useful in a biotechnological setting. While the observed hydrocarbon yields are low, this is to be expected in the absence of further engineering, and this work is a first step in development of *R. toruloides* as a platform for industrial production of hydrocarbons.

6 Concluding remarks

Over the last 5 to 10 years reduction in the price of crude oil and the emergence of problems associated with production of first generation biofuels (Searchinger *et al.*, 2008, Naik *et al.*, 2010) has led to a reduction in emphasis on biofuels, for example the scope of EU mandated target for the replacement of petroleum derived fuels with biofuels by 2020 was reduced from 20 % to 10 % (European Commission, 2009). However continued climate change means action on carbon emissions is more urgent than ever, and advanced microbial biofuels have the potential to displace emissions from the combustion of fossil fuels.

R. toruloides is well suited to the metabolism of lignocellulosic biomass (Wiebe *et al.*, 2012, Qi *et al.*, 2014, Fei *et al.*, 2016), and capable of accumulating lipid to 76 % of its dry biomass, high culture densities and fast growth resulting in high lipid yields (Li *et al.*, 2006, Li *et al.*, 2007, Zhang *et al.*, 2016), making this a good candidate for production of advanced biofuels. However production of higher value, lipid-derived compounds will require genetic engineering, made difficult by the lack of molecular genetic tools available for manipulation of this yeast.

Before starting this work Liu *et al.*, (2013) demonstrated ATMT of *R. toruloides* CBS 349. Due to the significant differences between strains CBS 349 and our strain of interest (CBS 14), it was necessary to demonstrate transformation of the latter strain. Since demonstrating ATMT of strain CBS 14 other groups have also reported ATMT of this and other strains, selecting for transformants by growth with the antibiotics nourseothricin, bleomycin,

zeocin as well as G418 or hygromycin used here (Liu *et al.*, 2013, Lin *et al.*, 2014, Takahashi *et al.*, 2014, Fillet *et al.*, 2015). As well as ATMT, electroporation of *Rhodotorula gracilis* ATCC 26217 has been reported and from available sequence, this strain is identical to strain CBS 14 (Takahashi *et al.*, 2014). ATMT works with high efficiency, and is robust to transformation with large DNA fragments, but this protocol is slow, taking almost two weeks from purified plasmid to transformed yeast. Relative to ATMT, electroporation works with lower efficiency and can be is sensitive to the use of large DNA fragments (de Groot *et al.*, 1998, Liu *et al.*, 2013, Takahashi *et al.*, 2014), but this protocol is faster, taking four days from purified plasmid to transformed yeast. During the course of this work all transformations were performed using ATMT, however it may be sensible to use electroporation for transformation of short DNA and ATMT for transformation of larger fragments or where higher efficiency is necessary.

The high GC content of *R. toruloides* DNA and constructs codon optimised for this yeast initially made routine cloning difficult. While being relatively slow, in-yeast assembly proved robust to assembly of constructs for targeted gene deletion, promoter screening and hydrocarbon biosynthesis despite the challenging nature of the plasmids constructed, for example plasmids for alkane biosynthesis were over 23 kb in size. Cut-sites included in vectors would allow classical restriction cloning or other *in vitro* methodologies to be used, which may be more appropriate for in insertion of small, simple fragments as these methodologies have the potential to be less time-consuming than in-yeast assembly.

Four inducible promoters were identified and characterised, each with independent induction and repression conditions making each potentially useful in different situations. Furthermore, they are also all independent of the *DAO1* inducible promoter (Liu *et al.*, 2015b). The *DAO1* promoter was identified and characterised in *R. toruloides* strain CBS 349; given the apparent differential regulation observed with the *MET16* and *CTR3* promoters between strains CBS 14 and the CBS 349-derived strain NCYC 1585, it cannot be guaranteed that this promoter would work as expected in *R. toruloides* CBS 14. However, assuming the *DAO1* promoter behaves the same in both strains, of the five inducible promoters described for *R. toruloides*, when studying gene function the *NAR1* promoter would be the first choice. The *NAR1* promoter showed both high induced activity and exceptionally tight regulation with the EGFP marker, and was the only promoter where conditional rescue of the *leu2* strain was observed. The *NAR1* promoter was not directly compared against the *DAO1* promoter, therefore it cannot be conclusively be determined which is more effective, however under repression conditions, silencing of the luciferase marker used to test the *DAO1* promoter was incomplete (Liu *et al.*, 2015b).

Production of heptadecane was observed after expression of *S. elongatus* fatty acyl-ACP reductase and fatty aldehyde decarbonylase, along with *E. coli* ferredoxin and ferredoxin reductase. The small alkane yield relative to the potential lipid available could be considered disappointing, however, the observed production was consistent with observations in *S. cerevisiae* (Buijs *et al.*, 2015). The low yield and exclusive production of heptadecane could be explained by the fact that only fatty acids conjugated to the FAS ACP domain

during their production are usable by fatty acyl-ACP reductase and that fatty acids shorter than 18 carbons in length are enclosed with the MPT domain of the FAS and therefore inaccessible to fatty acyl-ACP reductase, blocking production of shorter alkanes (Schirmer *et al.*, 2010, Fischer *et al.*, 2015).

Whilst the results of these hydrocarbon biosynthesis experiments are a first step towards production of hydrocarbons in *R. toruloides*, they demonstrate a requirement for further manipulation of this yeast for it to become a viable platform for production of microbial biofuels, and therefore a requirement for further tool development to achieve the required manipulation.

I was unable to demonstrate functional expression of fatty acyl-CoA reductase or carboxylic acid reductase, however use of these enzymes, and therefore fatty acyl-CoA or FFAs as a source for alkane production could increase potential yields. Were I able to express these enzymes, for optimal activity they would likely require targeting to the endoplasmic reticulum (Xu *et al.*, 2016). Whilst there is no reason to suggest subcellular protein targeting would be any different in *R. toruloides* relative to other yeasts, and that similar tags and epitopes could be used for protein targeting, this would require confirmation.

Another limit on production of alkanes form fatty aldehydes is oxidation of accumulated fatty aldehyde. After expression of this pathway in *S. cerevisiae*, alkane production could be increased almost 10-fold by deletion of hexadecenal dehydrogenase (*Hfd1*) (Buijs *et al.*, 2015) and further increased by disruption of other promiscuous aldehyde reductases or alcohol dehydrogenases (Zhou *et al.*, 2016). Overcoming this issue in *R. toruloides* would require a reproducible protocol for targeted gene deletion/disruption.

Unfortunately, during this work I was unable to demonstrate homologous integration and therefore targeted gene deletion/disruption, in agreement with results from other groups studying closely related strains (Lin *et al.*, 2014, Takahashi *et al.*, 2014). Using *R. toruloides* CBS 349, Koh *et al*. (2014) described targeted integration at the *CrtY* locus and generation of a *Ku70* mutant strain in which efficient homologous recombination could be performed using a protocol equivalent to that described in chapter 3. Assuming targeted integration cannot be achieved in *R. toruloides* CBS 14 and the work of Koh *et al*. (2014) can be reproduced in *R. toruloides* CBS 349, it may be pragmatic to focus future study on strain CBS 349, however there are techniques which have been developed which could facilitate targeted integration in strain CBS 14. CRISPR/Cas9 has been used to catalyse mutation and targeted integration in diverse eukaryotes including plants, animals, protists and fungi. Many variations on CRIPSR/Cas9 have been developed (Mali *et al.*, 2013) and many solutions to problems encountered, such as defined expression of guide RNA, are available. However, while development of a CRISPR/Cas9 for use in *R. toruloides* is feasible with existing molecular genetic tools, no counterselectable markers or extranuclear plasmid vectors are available for use in *R. toruloides*; a counterselectable plasmid would be useful to facilitate transient maintenance of genes for guide RNAs, and therefore simplify sequential deletion of multiple genes by CRISPR/Cas9.

One key lesson learned from this work is the importance of strain selection. At the start of this project it was decided to use strain CBS 14, most notably as this was a sequenced haploid strain available from strain collections (see section 1.2.1 for full discussion why), however over the intervening four years

genomes have been published for other strains including CBS 349. Other groups have demonstrated advantages of strain CBS 349 relative to CBS 14, most significantly targeted gene deletion has proved much easier in strain CBS 349 (Koh *et al*., 2014). Engineering for increased lipid production has also proved easier in strain CBS 349 (Zhang *et al.*, 2016). Were one starting work with a new organism it might be appropriate to work with different, related strains in parallel as differences, not initially apparent between strains may offer solutions to problems encountered later, however this would have to be weighed against the extra work and cost this would necessitate.

Another key lesson is the importance of different tools for the same job. For example, many different cloning technologies have been developed which apparently achieve the same ends in the assembly of plasmids, however different techniques have different advantages and disadvantages; Gibson assembly is a quick and easy technique for assembly of simple plasmids, but had low efficiency for assembly of the large, GC-rich plasmids required in this work, whereas in-yeast assembly is much slower and more labour intensive but demonstrated high efficiency, even for assembly of the large, GC-rich plasmids used in this work. More mature molecular genetic toolboxes for manipulation of other yeasts have allowed more extensive manipulations, and thus greater hydrocarbon yields in other yeasts including *Y. lipolytica* (Xu *et al.*, 2016) and even *S. cerevisiae* (Buijs *et al.*, 2015, Zhou *et al.*, 2016). However since the publication of a reproducible protocol for transformation of *R. toruloides* (Liu *et al.*, 2013), publications taking advantage of the potential of this yeast and modifying it for use in an economic setting are beginning to appear (Fillet *et al.*, 2015, Lee *et al.*, 2016, Zhang *et al.*, 2016). With

engineering, *R. toruloides* could likely outperform other, more developed yeasts for production of hydrocarbons and other fatty acid derived molecules and, whilst the tools and protocols developed in this work (Johns *et al.*, 2016) and others (Koh *et al.*, 2014, Lin *et al.*, 2014, Takahashi *et al.*, 2014, Liu *et al.*, 2015b, Wang *et al.*, 2016) will play a part in the domestication of this yeast, further work is needed.

References

Abbott EP, Ianiri G, Castoria R & Idnurm A (2013) Overcoming recalcitrant transformation and gene manipulation in Pucciniomycotina yeasts. *Appl Microbiol Biotechnol* **97**: 283-295.

Abe K, Fukui S & Kusanagi A (1984) Estimation of time lengths of cell-cycle phases from asynchronous cultures of the basidiomycetous yeast *Rhodosporidium toruloides*. *J Gen Appl Microbiol* **30**: 257-265.

Abramoff MD, Magalhaes PJ & Ram SJ (2004) Image Processing with ImageJ. *Biophotonics* **11**: 36-42.

Akhtar MK, Turner NJ & Jones PR (2013) Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proc Natl Acad Sci U S A* **110**: 87-92.

Al Hafid N & Christodoulou J (2015) Phenylketonuria: a review of current and future treatments. *Transl Pediatr* **4**: 304-317.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW & Noble WS (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202-208.

Bailey TL, Johnson J, Grant CE & Noble SW (2015) The MEME Suite. *Nucleic Acids Res* **43** 39-49.

Banks GR, Shelton PA, Kanuga N, Holden DW & Spanos A (1993) The *Ustilago maydis NAR1* gene encoding nitrate reductase activity: sequence and transcriptional regulation. *Gene* **131**: 69-78.

Banno I (1967) Studies on sexuality of *Rhodotorula*. *J Gen Appl Microbiol* **13**: 167-196.

Barbella R, Ciajolo A, Danna A & Bertoli C (1989) Effect of fuel aromaticity on diesel emissions. *Combust Flame* **77**: 267-277.

Barth G (1985) Genetic-regulation of isocitrate lyase in the yeast *Yarrowia lipolytica*. *Curr Genet* **10**: 119-124.

Becherel OJ, Jakob B, Cherry AL*, et al.* (2010) CK2 phosphorylation-dependent interaction between aprataxin and MDC1 in the DNA damage response. *Nucleic Acids Res* **38**: 1489-1503.

Beller HR, Goh EB & Keasling JD (2010) Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. *Appl Environ Microbiol* **76**: 1212-1223.

Bevan MW, Flavell RB & Chilton MD (1983) A chimaeric antibiotic-resistance gene as a selectable marker for plant-cell transformation. *Nature* **304**: 184-187.

Boulton CA & Ratledge C (1981a) Correlation of lipid-accumulation in yeasts with possession of ATP-citrate lyase. *J Gen Microbiol* **127**: 169-176.

Boulton CA & Ratledge C (1981b) ATP - citrate lyase - the regulatory enzyme for lipid biosynthesis in *Lipomyces starkeyi*. *J Gen Microbiol* **127**: 423-426.

Brachman CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P & Boeke JD (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**: 115-132.

Branzei D & Foiani M (2008) Regulation of DNA repair throughout the cell cycle. *Nat Rev Mol Cell Biol* **9**: 297-308.

Brown SR (2016) A design of experiments approach for engineering carbon metabolisim in the yeast *Saccharomyces cerevisiae*. Thesis, University of Exeter, Exeter UK.

Buck JW & Andrews JH (1999) Attachment of the yeast *Rhodosporidium toruloides* is mediated by adhesives localized at sites of bud cell development. *Appl Environ Microbiol* **65**: 465-471.

Buijs NA, Zhou YJ, Siewers V & Nielsen J (2015) Long-chain alkane production by the yeast *Saccharomyces cerevisiae*. *Biotechnol Bioeng* **112**: 1275-1279.

Bundock P, den Dulk-Ras A, Beijersbergen A & Hooykaas PJ (1995) Trans-kingdom T-DNA transfer from *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*. *EMBO J* **14**: 3206-3214.

Buzzini P, Innocenti M, Turchetti B, Libkind D, van Broock M & Mulinacci N (2007) Carotenoid profiles of yeasts belonging to the genera *Rhodotorula*, *Rhodosporidium*, *Sporobolomyces*, and *Sporidiobolus*. *Can J Microbiol* **53**: 1024-1031.

Cangelosi GA, Ankenbauer RG & Nester EW (1990) Sugars induce the *Agrobacterium* virulence genes through a periplasmic binding protein and a transmembrane signal protein. *Proc Natl Acad Sci U S A* **87**: 6708-6712.

Caplan A, Herrera-Estrella L, Inze D, Van Haute E, Van Montagu M, Schell J & Zambryski P (1983) Introduction of genetic material into plant cells. *Science* **222**: 815-821.

Carvalho ND, Arentshorst M, Jin Kwon M, Meyer V & Ram AF (2010) Expanding the ku70 toolbox for filamentous fungi: establishment of complementation vectors and recipient strains for advanced gene analyses. *Appl Microbiol Biotechnol* **87**: 1463-1473.

ChandGoyal T & Spotts RA (1996) Control of postharvest pear diseases using natural saprophytic yeast colonists and their combination with a low dosage of thiabendazole. *Postharvest Biol Tec* **7**: 51-64.

Chen B, Lee DY & Chang MW (2015) Combinatorial metabolic engineering of *Saccharomyces cerevisiae* for terminal alkene production. *Metab Eng* **31**: 53-61.

Chen X, Stone M, Schlagnhaufer C & Romaine CP (2000) A fruiting body tissue method for efficient *Agrobacterium*-mediated transformation of *Agaricus bisporus. Appl Environ Microbiol* **66** 4510–4513.

Chilton MD, Drummond MH, Merlo DJ & Sciaky D (1978) Highly conserved DNA of Ti plasmids overlaps T-DNA maintained in plant tumors. *Nature* **275**: 147-149.

Chilton MD, Drummond MH, Merio DJ, Sciaky D, Montoya AL, Gordon MP & Nester EW (1977) Stable incorporation of plasmid DNA into higher plant cells: the molecular basis of crown gall tumorigenesis. *Cell* **11**: 263-271.

Christie PJ (1997) *Agrobacterium tumefaciens* T-complex transport apparatus: a paradigm for a new family of multifunctional transporters in eubacteria. *J Bacteriol* **179**: 3085-3094.

Claros MG & Vincens P (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. *Eur J Biochem* **241**: 779-786.

Coelho MA, Almeida JM, Hittinger CT & Goncalves P (2015) Draft genome sequence of *Sporidiobolus salmonicolor* CBS 6832, a red-pigmented basidiomycetous yeast. *Genome Announc* **3**.

Couturier M, Navarro D, Olive C, Chevret D, Haon M, Favel A, Lesage-Meessen L, Henrissat B, Coutinho PM & Berrin JG (2012) Post-genomic analyses of fungal lignocellulosic biomass degradation reveal the unexpected potential of the plant pathogen *Ustilago maydis*. *BMC Genomics* **13**: 57.

Critchlow SE & Jackson SP (1998) DNA end-joining: from yeast to man. *Trends Biochem Sci* **23**: 394-398.

da Silva TL, Gouveia L & Reis A (2014) Integrated microbial processes for biofuels and high value-added products: the way to improve the cost effectiveness of biofuel production. *Appl Microbiol Biotechnol* **98**: 1043-1053.

Dayananda C, Sarada R, Rani MU, Shamala TR & Ravishankar GA (2007) Autotrophic cultivation of *Botryococcus braunii* for the production of hydrocarbons and exopolysaccharides in various media. *Biomass Bioenerg* **31**: 87-93.

de Groot MJ, Bundock P, Hooykaas PJ & Beijersbergen AG (1998) *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. *Nat Biotechnol* **16**: 839-842.

de Jong B, Siewers V & Nielsen J (2012) Systems biology of yeast: enabling technology for development of cell factories for production of advanced biofuels. *Curr Opin Biotechnol* **23**: 624-630.

Dereeper A, Guignon V, Blanc G*, et al.* (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* **36**: W465-469.

Doshi R, Nguyen T & Chang G (2013) Transporter-mediated biofuel secretion. *Proc Natl Acad Sci U S A* **110**: 7642-7647.

Dujon B, Sherman D, Fischer G*, et al.* (2004) Genome evolution in yeasts. *Nature* **430**: 35-44.

Duplessis S, Cuomo CA, Lin YC*, et al.* (2011) Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci U S A* **108**: 9166-9171.

Eisenberg T & Buttner S (2014) Lipids and cell death in yeast. *Fems Yeast Res* **14**: 179-197.

Ernst JF & Chan RK (1985) Characterization of *Saccharomyces cerevisiae* mutants supersensitive to aminoglycoside antibiotics. *J Bacteriol* **163**: 8-14.

European Commission (2009) Directive 2009/28/EC of the European Parliament and of the Council of 23rd April 2009. *Official Journal of the European Union* 16-62.

Evans CT & Ratledge C (1984a) Influence of nitrogen-metabolism on lipid-accumulation by *Rhodosporidium toruloides* Cbs-14. *J Gen Microbiol* **130**: 1705-1710.

Evans CT & Ratledge C (1984b) Effect of nitrogen-source on lipid-accumulation in oleaginous yeasts. *J Gen Microbiol* **130**: 1693-1704.

Fazal MA, Haseeb ASMA & Masjuki HH (2011) Biodiesel feasibility study: An evaluation of material compatibility; performance; emission and engine durability. *Renew Sust Energ Rev* **15**: 1314-1324.

Fei Q, O'Brien M, Nelson R, Chen X, Lowell A & Dowe N (2016) Enhanced lipid production by *Rhodosporidium toruloides* using different fed-batch feeding strategies with lignocellulosic hydrolysate as the sole carbon source. *Biotechnol Biofuels* **9**: 130.

Fernandez-Lafuente R (2010) Lipase from *Thermomyces lanuginosus*: Uses and prospects as an industrial biocatalyst. *J Mol Catal B-Enzym* **62**: 197-212.

Fillet S, Gibert J, Suarez B, Lara A, Ronchel C & Adrio JL (2015) Fatty alcohols production by oleaginous yeast. *J Ind Microbiol Biot* **42**: 1463-1472.

Filonow AB, Vishniac HS, Anderson JA & Janisiewicz WJ (1996) Biological control of *Botrytis cinerea* in apple by yeasts from various habitats and their putative mechanisms of antagonism. *Biol Control* **7**: 212-220.

Firrincieli A, Otillar R, Salamov A, Schmutz J, Khan Z, Redman RS, Fleck ND, Lindquist E, Grigoriev IV & Doty SL (2015) Genome sequence of the plant growth promoting endophytic yeast *Rhodotorula graminis* WP1. *Front Microbiol* **6**: 978.

Fischer M, Rhinow D, Zhu Z, Mills DJ, Zhao ZK, Vonck J & Grininger M (2015) Cryo-EM structure of fatty acid synthase (FAS) from *Rhodosporidium toruloides* provides insights into the evolutionary development of fungal FAS. *Protein Sci* **24**: 987-995.

Fortman JL, Chhabra S, Mukhopadhyay A, Chou H, Lee TS, Steen E & Keasling JD (2008) Biofuel alternatives to ethanol: pumping the microbial well. *Trends Biotechnol* **26**: 375-381.

Fraaije BA, Bayon C, Atkins S, Cools HJ, Lucas JA & Fraaije MW (2012) Risk assessment studies on succinate dehydrogenase inhibitors, the new weapons in the battle to control *Septoria* leaf blotch in wheat. *Mol Plant Pathol* **13**: 263-275.

Fraley RT, Rogers SG, Horsch RB*, et al.* (1983) Expression of bacterial genes in plant-cells. *P Natl Acad Sci-Biol* **80**: 4803-4807.

Fritz RR, Hodgins DS & Abell CW (1976) Phenylalanine ammonia-lyase. Induction and purification from yeast and clearance in mammals. *J Biol Chem* **251**: 4646-4650.

Fullner KJ & Nester EW (1996) Temperature affects the T-DNA transfer machinery of *Agrobacterium tumefaciens*. *J Bacteriol* **178**: 1498-1504.

Gabriel F, Accoceberry I, Bessoule JJ, Salin B, Lucas-Guerin M, Manon S, Dementhon K & Noel T (2014) A Fox2-dependent fatty acid beta-oxidation pathway coexists both in peroxisomes and mitochondria of the ascomycete yeast *Candida lusitaniae*. *Plos One* **9**.

Gamez A, Sarkissian CN, Wang L*, et al.* (2005) Development of pegylated forms of recombinant *Rhodosporidium toruloides* phenylalanine ammonia-lyase for the treatment of classical phenylketonuria. *Mol Ther* **11**: 986-989.

Garay LA, Boundy-Mills KL & German JB (2014) Accumulation of high-value lipids in single-cell microorganisms: A mechanistic approach and future perspectives. *J Agr Food Chem* **62**: 2709-2727.

Gatz C & Quail PH (1988) Tn10-encoded tet repressor can regulate an operator-containing plant promoter. *Proc Natl Acad Sci U S A* **85**: 1394-1397.

Gebhart D, Bahrami AK & Sil A (2006) Identification of a copper-inducible promoter for use in ectopic expression in the fungal pathogen *Histoplasma capsulatum*. *Eukaryot Cell* **5**: 935-944.

Gelvin SB (2003) *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol Mol Biol Rev* **67**: 16-37, table of contents.

Gelvin SB (2006) Agrobacterium virulence gene induction. *Methods Mol Biol* **343**: 77-84.

Gelvin SB (2010) Plant proteins involved in *Agrobacterium*-mediated genetic transformation. *Annu Rev Phytopathol* **48**: 45-68.

Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, 3rd & Smith HO (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**: 343-345.

Gilbert HJ & Tully M (1982) Synthesis and degradation of phenylalanine ammonia-lyase of *Rhodosporidium toruloides*. *J Bacteriol* **150**: 498-505.

Gilbert HJ, Stephenson JR & Tully M (1983) Control of synthesis of functional mRNA coding for phenylalanine ammonia-lyase from *Rhodosporidium toruloides*. *J Bacteriol* **153**: 1147-1154.

Gilbert HJ, Clarke IN, Gibson RK, Stephenson JR & Tully M (1985) Molecular cloning of the phenylalanine ammonia lyase gene from *Rhodosporidium toruloides* in *Escherichia coli* K-12. *J Bacteriol* **161**: 314-320.

Goins CL, Gerik KJ & Lodge JK (2006) Improvements to gene deletion in the fungal pathogen *Cryptococcus neoformans*: absence of Ku proteins increases homologous recombination, and co-transformation of independent DNA molecules allows rapid complementation of deletion phenotypes. *Fungal Genet Biol* **43**: 531-544.

Gong ZW, Wang Q, Shen HW, Hu CM, Jin GJ & Zhao ZBK (2012) Co-fermentation of cellobiose and xylose by *Lipomyces starkeyi* for lipid production. *Bioresource Technol* **117**: 20-24.

Gorbunova V & Levy AA (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. *Nucleic Acids Res* **25**: 4650-4657.

Gossen M & Bujard H (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc Natl Acad Sci U S A* **89**: 5547-5551.

Gouda MK, Omar SH & Aouad LM (2008) Single cell oil production by *Gordonia sp.* DG using agro-industrial wastes. *World J Microb Biot* **24**: 1703-1711.

Grant CE, Bailey TL & Noble WS (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017-1018.

Grawunder U, Wilm M, Wu X, Kulesza P, Wilson TE, Mann M & Lieber MR (1997) Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature* **388**: 492-495.

Grigoriev IV, Nikitin R, Haridas S*, et al.* (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res* **42**: D699-704.

Gupta P & Phulara SC (2015) Metabolic engineering for isoprenoid-based biofuel production. *J Appl Microbiol* **119**: 605-619.

Gupta S, Mana-Capelli S, McLean JR, Chen CT, Ray S, Gould KL & McCollum D (2013) Identification of SIN pathway targets reveals mechanisms of crosstalk between NDR kinase pathways. *Curr Biol* **23**: 333-338.

Haas MJ, McAloon AJ, Yee WC & Foglia TA (2006) A process model to estimate biodiesel production costs. *Bioresour Technol* **97**: 671-678.

Hausmann A & Sandmann G (2000) A single five-step desaturase is involved in the carotenoid biosynthesis pathway to beta-carotene and torulene in *Neurospora crassa*. *Fungal Genet Biol* **30**: 147-153.

Hawksworth DL (2011) A new dawn for the naming of fungi: impacts of decisions made in Melbourne in July 2011 on the future publication and regulation of fungal names. *IMA Fungus* **2**: 155-162.

Herman NA & Zhang W (2016) Enzymes for fatty acid-based hydrocarbon biosynthesis. *Curr Opin Chem Biol* **35**: 22-28.

Herrera-Estrella A, Van Montagu M & Wang K (1990) A bacterial peptide acting as a plant nuclear targeting signal: the amino-terminal portion of *Agrobacterium* VirD2 protein directs a beta-galactosidase fusion protein into tobacco nuclei. *Proc Natl Acad Sci U S A* **87**: 9534-9537.

Herrera-Estrella A, Chen ZM, Van Montagu M & Wang K (1988) VirD proteins of *Agrobacterium tumefaciens* are required for the formation of a covalent DNA--protein complex at the 5' terminus of T-strand molecules. *EMBO J* **7**: 4055-4062.

Hoekema A, Hirsch PR, Hooykaas PJJ & Schilperoort RA (1983) A binary plant vector strategy based on separation of Vir-region and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. *Nature* **303**: 179-180.

Holsters M, de Waele D, Depicker A, Messens E, van Montagu M & Schell J (1978) Transfection and transformation of *Agrobacterium tumefaciens*. *Mol Gen Genet* **163**: 181-187.

Howard EA, Zupan JR, Citovsky V & Zambryski PC (1992) The VirD2 protein of A. tumefaciens contains a C-terminal bipartite nuclear localization signal: implications for nuclear uptake of DNA in plant cells. *Cell* **68**: 109-118.

Howard TP, Middelhaufe S, Moore K*, et al.* (2013) Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in *Escherichia coli*. *Proc Natl Acad Sci U S A* **110**: 7636-7641.

Hu J, Li B & Kihara D (2005) Limitations and potentials of current motif discovery algorithms. *Nucleic Acids Res* **33** 4899-4913.

Hu MC & Davidson N (1987) The inducible lac operator-repressor system is functional in mammalian cells. *Cell* **48**: 555-566.

Huang C, Chen XF, Xiong L, Chen XD & Ma LL (2012) Oil production by the yeast *Trichosporon dermatis* cultured in enzymatic hydrolysates of corncobs. *Bioresour Technol* **110**: 711-714.

Huang XF, Liu JN, Lu LJ, Peng KM, Yang GX & Liu J (2016) Culture strategies for lipid production using acetic acid as sole carbon source by *Rhodosporidium toruloides*. *Bioresour Technol* **206**: 141-149.

Isikgor FH & Becer CR (2015) Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers. *Polym Chem-Uk* **6**: 4497-4559.

Jacob Z (1992) Yeast lipids - extraction, quality analysis, and acceptability. *Crit Rev Biotechnol* **12**: 463-491.

Jenni S, Leibundgut M, Boehringer D, Frick C, Mikolasek B & Ban N (2007) Structure of fungal fatty acid synthase and implications for iterative substrate shuttling. *Science* **316**: 254-261.

Jiang C & Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* **10**: 161-172.

Johns AM, Love J & Aves SJ (2016) Four Inducible Promoters for Controlled Gene Expression in the Oleaginous Yeast Rhodotorula toruloides. *Front Microbiol* **7**: 1666.

Johnston M & Davis RW (1984) Sequences that regulate the divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. *Mol Cell Biol* **4**: 1440-1448.

Jonsson JJ, Foresman MD, Wilson N & Mcivor RS (1992) Intron Requirement for Expression of the Human Purine Nucleoside Phosphorylase Gene. *Nucleic Acids Research* **20**: 3191-3198.

Kalscheuer R, Luftmann H & Steinbuchel A (2004) Synthesis of novel lipids in *Saccharomyces cerevisiae* by heterologous expression of an unspecific bacterial acyltransferase. *Appl Environ Microb* **70**: 7119-7125.

Kamisaka Y, Kimura K, Uemura H & Yamaoka M (2013) Overexpression of the active diacylglycerol acyltransferase variant transforms *Saccharomyces cerevisiae* into an oleaginous yeast. *Appl Microbiol Biotechnol* **97**: 7345-7355.

Kampman B, Verbeek R, van Grinsven A, van Mensch P, Crozen H & Patuleia A (2013) Options to increase EU biofuels volumes beyond the curent blending limits. p.^pp.

Kanehisa M & Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**: 27-30.

Kanehisa M, Sato Y, Kawashima M, Furumichi M & Tanabe M (2016) KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**: D457-462.

Kang DH, Hyeon JE, You SK, Kim SW & Han SO (2014) Efficient enzymatic degradation process for hydrolysis activity of the Carrageenan from red algae in marine biomass. *J Biotechnol* **192 Pt A**: 108-113.

Kilaru S & Steinberg G (2015) Yeast recombination-based cloning as an efficient way of constructing vectors for *Zymoseptoria tritici*. *Fungal Genet Biol* **79**: 76-83.

Kim SI, Veena & Gelvin SB (2007) Genome-wide analysis of *Agrobacterium* T-DNA integration sites in the *Arabidopsis* genome generated under non-selective conditions. *Plant J* **51**: 779-791.

Kleinboelting N, Huep G, Appelhagen I, Viehoever P, Li Y & Weisshaar B (2015) The structural features of thousands of T-DNA insertion sites are consistent with a double-strand break repair-based insertion mechanism. *Mol Plant* **8**: 1651-1664.

Kleinegris DM, van Es MA, Janssen M, Brandenburg WA & Wijffels RH (2010) Carotenoid fluorescence in *Dunaliella salina*. *J Appl Phycol* **22**: 645-649.

Koh CMJ, Liu YB, Moehninsi, Du MG & Ji LH (2014) Molecular characterization of *KU70* and *KU80* homologues and exploitation of a *ku70*-deficient mutant for improving gene deletion frequency in *Rhodosporidium toruloides*. *Bmc Microbiol* **14**.

Kojic M & Holloman WK (2000) Shuttle vectors for genetic manipulations in *Ustilago maydis*. *Can J Microbiol* **46**: 333-338.

Koukolikova-Nicola Z & Hohn B (1993) How does the T-DNA of *Agrobacterium tumefaciens* find its way into the plant cell nucleus? *Biochimie* **75**: 635-638.

Krappmann S (2007) Gene targeting in filamentous fungi: the benefits of impaired repair. *Fungal Biology Reviews* **21**: 25-29.

Krappmann S, Sasse C & Braus GH (2006) Gene targeting in *Aspergillus fumigatus* by homologous recombination is facilitated in a nonhomologous end- joining-deficient genetic background. *Eukaryot Cell* **5**: 212-215.

Kuck U & Hoff B (2010) New tools for the genetic manipulation of filamentous fungi. *Appl Microbiol Biotechnol* **86**: 51-62.

Kumar P, Barrett DM, Delwiche MJ & Stroeve P (2009) Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production. *Ind Eng Chem Res* **48**: 3713-3729.

Kumar S, Gupta N & Pakshirajan K (2015) Simultaneous lipid production and dairy wastewater treatment using *Rhodococcus opacus* in a batch bioreactor for potential biodiesel application. *Journal of Environmental Chemical Engineering* **3**: 1630-1636.

Kumar S, Kushwaha H, Bachhawat AK, Raghava GP & Ganesan K (2012) Genome sequence of the oleaginous red yeast *Rhodosporidium toruloides* MTCC 457. *Eukaryot Cell* **11**: 1083-1084.

Kunik T, Tzfira T, Kapulnik Y, Gafni Y, Dingwall C & Citovsky V (2001) Genetic transformation of HeLa cells by *Agrobacterium*. *Proc Natl Acad Sci U S A* **98**: 1871-1876.

Kuo CY, Chou SY & Huang CT (2004) Cloning of glyceraldehyde-3-phosphate dehydrogenase gene and use of the gpd promoter for transformation in *Flammulina velutipes*. *Appl Microbiol Biotechnol* **65**: 593-599.

Labbe S & Thiele DJ (1999) Copper ion inducible and repressible promoter systems in yeast. *Methods Enzymol* **306**: 145-153.

Lantermann AB, Straub T, Stralfors A, Yuan GC, Ekwall K & Korber P (2010) Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nat Struct Mol Biol* **17**: 251-U215.

Lazo GR, Stein PA & Ludwig RA (1991) A DNA transformation-competent *Arabidopsis* genomic library in *Agrobacterium*. *Biotechnology (N Y)* **9**: 963-967.

Lee JJ, Chen L, Cao B & Chen WN (2016) Engineering *Rhodosporidium toruloides* with a membrane transporter facilitates production and separation of carotenoids and lipids in a bi-phasic culture. *Appl Microbiol Biotechnol* **100**: 869-877.

Lee JJ, Chen L, Shi J, Trzcinski A & Chen WN (2014) Metabolomic profiling of *Rhodosporidium toruloides* grown on glycerol for carotenoid production during different growth phases. *J Agric Food Chem* **62**: 10203-10209.

Lee SY, Kim HM & Cheon S (2015) Metabolic engineering for the production of hydrocarbon fuels. *Curr Opin Biotechnol* **33**: 15-22.

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR & Nislow C (2007) A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* **39**: 1235-1244.

Leroux P & Berthier G (1988) Resistance to carboxin and fenfuram in *Ustilago nuda* (Jens) Rostr, the causal agent of barley loose smut. *Crop Prot* **7**: 16-19.

Lewandowski I, Clifton-Brown JC, Scurlock JMO & Huisman W (2000) *Miscanthus*: European experience with a novel energy crop. *Biomass Bioenerg* **19**: 209-227.

Li Q, Du W & Liu DH (2008) Perspectives of microbial oils for biodiesel production. *Appl Microbiol Biot* **80**: 749-756.

Li YH, Zhao ZB & Bai FW (2007) High-density cultivation of oleaginous yeast *Rhodosporidium toruloides* Y4 in fed-batch culture. *Enzyme Microb Tech* **41**: 312-317.

Li YH, Liu B, Zhao ZB & Bai FW (2006) Optimized culture medium and fermentation conditions for lipid production by *Rhodosporidium toruloides*. *Sheng Wu Gong Cheng Xue Bao* **22**: 650-656.

Liao JC, Mi L, Pontrelli S & Luo S (2016) Fuelling the future: microbial engineering for the production of sustainable biofuels. *Nat Rev Microbiol* **14**: 288-304.

Lin X, Yang F, Zhou Y, Zhu Z, Jin G, Zhang S & Zhao ZK (2012) Highly-efficient colony PCR method for red yeasts and its application to identify mutations within two leucine auxotroph mutants. *Yeast* **29**: 467-474.

Lin XP, Wang YN, Zhang SF, Zhu ZW, Zhou YJJ, Yang F, Sun WY, Wang XY & Zhao ZBK (2014) Functional integration of multiple genes into the genome of the oleaginous yeast *Rhodosporidium toruloides*. *Fems Yeast Res* **14**: 547-555.

Litchfield DW (2003) Protein kinase CK2: structure, regulation and role in cellular decisions of life and death. *Biochem J* **369**: 1-15.

Liu H, Zhao X, Wang F, Li Y, Jiang X, Ye M, Zhao ZK & Zou H (2009) Comparative proteomic analysis of *Rhodosporidium toruloides* during lipid accumulation. *Yeast* **26**: 553-566.

Liu L, Pan A, Spofford C, Zhou N & Alper HS (2015a) An evolutionary metabolic engineering approach for enhancing lipogenesis in *Yarrowia lipolytica*. *Metab Eng* **29**: 36-45.

Liu R, Zhu F, Lu L, Fu A, Lu J, Deng Z & Liu T (2014a) Metabolic engineering of fatty acyl-ACP reductase-dependent pathway to improve fatty alcohol production in *Escherichia coli*. *Metab Eng* **22**: 10-21.

Liu Y, Koh CM, Ngoh ST & Ji L (2015b) Engineering an efficient and tight D-amino acid-inducible gene expression system in *Rhodosporidium*/*Rhodotorula* species. *Microb Cell Fact* **14**: 170.

Liu Y, Koh CM, Sun L, Hlaing MM, Du M, Peng N & Ji L (2013) Characterization of glyceraldehyde-3-phosphate dehydrogenase gene *RtGPD1* and development of genetic transformation method by dominant selection in oleaginous yeast *Rhodosporidium toruloides*. *Appl Microbiol Biotechnol* **97**: 719-729.

Liu Y, Wang C, Yan J, Zhang W, Guan W, Lu X & Li S (2014b) Hydrogen peroxide-independent production of alpha-alkenes by OleTJE P450 fatty acid decarboxylase. *Biotechnol Biofuels* **7**: 28.

Ljungdahl PO & Daignan-Fornier B (2012) Regulation of Amino Acid, Nucleotide, and Phosphate Metabolism in Saccharomyces cerevisiae. *Genetics* **190**: 885-929.

Lu HP, Lu LF, Zeng LZ, Fu D, Xiang HL, Yu T & Zheng XD (2014) Effect of chitin on the antagonistic activity of *Rhodosporidium paludigenum* against Penicillium expansum in apple fruit. *Postharvest Biol Tec* **92**: 9-15.

Lugones LG, Scholtmeijer K, Klootwijk R & Wessels JG (1999) Introns are necessary for mRNA accumulation in Schizophyllum commune. *Mol Microbiol* **32**: 681-689.

Mali P, Esvelt KM & Church GM (2013) Cas9 as a versatile tool for engineering biology. *Nature Methods* **10**: 957-963.

Manetti AG, Rosetto M & Maundrell KG (1994) *nmt2* of fission yeast: a second thiamine-repressible gene co-ordinately regulated with *nmt1*. *Yeast* **10**: 1075-1082.

Mantis NJ & Winans SC (1992) The *Agrobacterium tumefaciens* vir gene transcriptional activator virG is transcriptionally induced by acid pH and other stress stimuli. *J Bacteriol* **174**: 1189-1196.

Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR & Ploegh HL (2015) Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol* **33**: 538-542.

Mascarenhas D, Mettler IJ, Pierce DA & Lowe HW (1990) Intron-mediated enhancement of heterologous gene expression in maize. *Plant Mol Biol* **15**: 913-920.

Mattioni T, Louvion JF & Picard D (1994) Regulation of protein activities by fusion to steroid binding domains. *Methods Cell Biol* **43 Pt A**: 335-352.

Matzke AJ & Matzke MA (1998) Position effects and epigenetic silencing of plant transgenes. *Curr Opin Plant Biol* **1**: 142-148.

Maundrell K (1990) *nmt1* of fission yeast. A highly transcribed gene completely repressed by thiamine. *J Biol Chem* **265**: 10857-10864.

Michielse CB, Hooykaas PJ, van den Hondel CA & Ram AF (2005) *Agrobacterium*-mediated transformation as a tool for functional genomics in fungi. *Curr Genet* **48**: 1-17.

Moharana TR, Byreddy AR, Puri M, Barrow C & Rao NM (2016) Selective enrichment of omega-3 fatty acids in oils by phospholipase A1. *Plos One* **11**.

Muniraj IK, Uthandi SK, Hu Z, Xiao L & Zhan X (2015) Microbial lipid production from renewable and waste materials for second-generation biodiesel feedstock. *Environmental Technology Reviews* **4**: 1-16.

Naik SN, Goud VV, Rout PK & Dalai AK (2010) Production of first and second generation biofuels: A comprehensive review. *Renew Sust Energ Rev* **14**: 578-597.

Nelson B, Kurischko C, Horecka J*, et al.* (2003) RAM: a conserved signaling network that regulates Ace2p transcriptional activity and polarized morphogenesis. *Mol Biol Cell* **14**: 3782-3803.

Neveu B, Belzile F & Belanger RR (2007) Cloning of the glyceraldehyde-3-phosphate dehydrogenase gene from *Pseudozyma flocculosa* and functionality of its promoter in two *Pseudozyma species*. *Antonie Van Leeuwenhoek* **92**: 245-255.

Nicaud J-M, Coq A-MC-L, Rossignol T & Morin N (2014) Protocols for monitoring growth and lipid accumulation in oleaginous yeasts. p.^pp. 1-17. Humana Press, Totowa, NJ.

Niklitschek M, Alcaino J, Barahona S*, et al.* (2008) Genomic organization of the structural genes controlling the astaxanthin biosynthesis pathway of *Xanthophyllomyces dendrorhous*. *Biol Res* **41**: 93-108.

Okunuki K (1931) Beiträge zur Kenntnis der rosafarbigen Sprosspilze. . *Journal of Japanese Botany* **5**: 285-322.

Ouyang SQ, Beecher CN, Wang K, Larive CK & Borkovich KA (2015) Metabolic impacts of using nitrogen and copper-regulated promoters to regulate gene expression in *Neurospora crassa*. *G3-Genes Genom Genet* **5**: 1899-1908.

Ovcharenko I, Loots GG, Hardison RC, Miller W & Stubbs L (2004) zPicture: dynamic alignment and visualization tool for analyzing conservation profiles. *Genome Res* **14**: 472-477.

Pan P & Hua Q (2012) Reconstruction and in silico analysis of metabolic network for an oleaginous yeast, *Yarrowia lipolytica*. *Plos One* **7**: e51535.

Patterson BW, Zhao G, Elias N, Hachey DL & Klein S (1999) Validation of a new procedure to determine plasma fatty acid concentration and isotopic enrichment. *J Lipid Res* **40**: 2118-2124.

Paul D, Magbanua Z, Arick M, 2nd, French T, Bridges SM, Burgess SC & Lawrence ML (2014) Genome sequence of the oleaginous yeast *Rhodotorula glutinis* ATCC 204091. *Genome Announc* **2**.

Perez-Garcia O, Escalante FM, de-Bashan LE & Bashan Y (2011) Heterotrophic cultures of microalgae: metabolism and potential products. *Water Res* **45**: 11-36.

Poirier Y, Antonenkov VD, Glumoff T & Hiltunen JK (2006) Peroxisomal beta-oxidation - a metabolic pathway with multiple functions. *Biochim Biophys Acta* **1763**: 1413-1426.

Puchta H (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J Exp Bot* **56**: 1-14.

Punt PJ, Dingemanse MA, Kuyvenhoven A, Soede RD, Pouwels PH & van den Hondel CA (1990) Functional elements in the promoter region of the *Aspergillus nidulans gpdA* gene encoding glyceraldehyde-3-phosphate dehydrogenase. *Gene* **93**: 101-109.

Qi F, Kitahara Y, Wang ZT, Zhao XB, Du W & Liu DH (2014) Novel mutant strains of *Rhodosporidium toruloides* by plasma mutagenesis approach and their tolerance for inhibitors in lignocellulosic hydrolyzate. *J Chem Technol Biot* **89**: 735-742.

Quan J & Tian J (2014) Circular polymerase extension cloning. *DNA Cloning and Assembly Methods,*(Valla S & Lale R, eds.), p.^pp. 103-117. Humana Press, Totowa, NJ.

Rakicka M, Lazar Z, Dulermo T, Fickers P & Nicaud JM (2015) Lipid production by the oleaginous yeast Yarrowia lipolytica using industrial by-products under different culture conditions. *Biotechnol Biofuels* **8**: 104.

Ratledge C (2004) Fatty acid biosynthesis in microorganisms being used for single cell oil production. *Biochimie* **86**: 807-815.

Ratledge C (2014) The role of malic enzyme as the provider of NADPH in oleaginous microorganisms: a reappraisal and unsolved problems. *Biotechnol Lett* **36**: 1557-1568.

Refaat AA (2009) Correlation between the chemical structure of biodiesel and its physical properties. *Int J Environ Sci Te* **6**: 677-694.

Rennerfelt E (1937) Undersökningar över svampinfektionen i slipmassa och dess utveckling däri. *Svenska Skogsvårdsföreningens Tidskrift* **35**: 43-159.

Rolland S, Jobic C, Fèvre M & Bruel C (2003) *Agrobacterium*-mediated transformation of *Botrytis cinerea*, simple purification of monokaryotic transformants and rapid conidia-based identification of the transfer-DNA host genomic DNA flanking sequences. *Curr Genet* **44** 164-171.

Rude MA, Baron TS, Brubaker S, Alibhai M, Del Cardayre SB & Schirmer A (2011) Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid decarboxylase from *Jeotgalicoccus* species. *Appl Environ Microbiol* **77**: 1718-1727.

Ruff JA, Lodge JK & Baker LG (2009) Three galactose inducible promoters for use in *C. neoformans* var. grubii. *Fungal Genet Biol* **46**: 9-16.

Ryu S, Hipp J & Trinh CT (2015) Activating and elucidating metabolism of complex sugars in *Yarrowia lipolytica*. *Appl Environ Microbiol* **82**: 1334-1345.

Sambrook J & Russell DW (2001) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Schirmer A, Rude MA, Li XZ, Popova E & del Cardayre SB (2010) Microbial biosynthesis of alkanes. *Science* **329**: 559-562.

Schweigert N, Zehnder AJ & Eggen RI (2001) Chemical properties of catechols and their molecular modes of toxic action in cells, from microorganisms to mammals. *Environ Microbiol* **3**: 81-91.

Searchinger T, Heimlich R, Houghton RA, Dong FX, Elobeid A, Fabiosa J, Tokgoz S, Hayes D & Yu TH (2008) Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science* **319**: 1238-1240.

Shapiro HM (2003) *Practical flow cytometry*. Wiley-Liss, Hoboken.

Shi S, Valle-Rodriguez JO, Siewers V & Nielsen J (2014) Engineering of chromosomal wax ester synthase integrated *Saccharomyces cerevisiae* mutants for improved biosynthesis of fatty acid ethyl esters. *Biotechnol Bioeng* **111**: 1740-1747.

Shima Y, Ito Y, Kaneko S, Hatabayashi H, Watanabe Y, Adachi Y & Yabe K (2009) Identification of three mutant loci conferring carboxin-resistance and development of a novel transformation system in *Aspergillus oryzae*. *Fungal Genet Biol* **46**: 67-76.

Shimoda N, Toyoda-Yamamoto A, Nagamine J, Usami S, Katayama M, Sakagami Y & Machida Y (1990) Control of expression of *Agrobacterium* vir genes by synergistic actions of phenolic signal molecules and monosaccharides. *Proc Natl Acad Sci U S A* **87**: 6684-6688.

Sidhu YS, Cairns TC, Chaudhari YK, Usher J, Talbot NJ, Studholme DJ, Csukai M & Haynes K (2015) Exploitation of sulfonylurea resistance marker and non-homologous end joining mutants for functional analysis in *Zymoseptoria tritici*. *Fungal Genet Biol* **79**: 102-109.

Siedenberg D, Mestric S, Ganzlin M, Schmidt M, Punt PJ, van den Hondel C & Rinas U (1999) *GlaA* promoter controlled production of a mutant green

fluorescent protein (S65T) by recombinant *Aspergillus niger* during growth on defined medium in batch and fed-batch cultures. *Biotechnol Prog* **15**: 43-50.

Singh MV & Weil PA (2002) A method for plasmid purification directly from yeast. *Anal Biochem* **307**: 13-17.

Smith EF & Townsend CO (1907) A plant-tumor of bacterial origin. *Science* **25**: 671-673.

Solow SP, Sengbusch J & Laird MW (2005) Heterologous protein production from the inducible *MET25* promoter in *Saccharomyces cerevisiae*. *Biotechnol Prog* **21**: 617-620.

Srivastava M, Nambiar M, Sharma S*, et al.* (2012) An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell* **151**: 1474-1487.

Stachel SE, Nester EW & Zambryski PC (1986) A plant cell factor induces *Agrobacterium tumefaciens* vir gene expression. *Proc Natl Acad Sci U S A* **83**: 379-383.

Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, Del Cardayre SB & Keasling JD (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463**: 559-562.

Sudesh K, Abe H & Doi Y (2000) Synthesis, structure and properties of polyhydroxyalkanoates: biological polyesters. *Prog Polym Sci* **25**: 1503-1555.

Takahashi S, Okada H, Abe K & Kera Y (2014) Genetic transformation of the yeast *Rhodotorula gracilis* ATCC 26217 by electroporation. *Appl Biochem Micro*+ **50**: 624-628.

Tanaka E, Bailey TL & Keich U (2014) Improving MEME via a two-tiered significance analysis. *Bioinformatics* **30** 1956-1973.

Tavares S, Ramos AP, Pires AS*, et al.* (2014) Genome size analyses of Pucciniales reveal the largest fungal genomes. *Front Plant Sci* **5**: 422.

Thiru M, Sankh S & Rangaswamy V (2011) Process for biodiesel production from *Cryptococcus curvatus*. *Bioresour Technol* **102**: 10436-10440.

Tinland B, Hohn B & Puchta H (1994) *Agrobacterium tumefaciens* transfers single-stranded transferred DNA (T-DNA) into the plant cell nucleus. *Proc Natl Acad Sci U S A* **91**: 8000-8004.

Toome M, Ohm RA, Riley RW*, et al.* (2014) Genome sequencing provides insight into the reproductive biology, nutritional mode and ploidy of the fern pathogen *Mixia osmundae*. *New Phytol* **202**: 554-564.

Tsigie YA, Wang CY, Truong CT & Ju YH (2011) Lipid production from *Yarrowia lipolytica* Po1g grown in sugarcane bagasse hydrolysate. *Bioresour Technol* **102**: 9216-9222.

Tully M (1985) Enrichment of mutants of *Rhodosporidium toruloides* by the use of inositol starvation. *J Basic Microb* **25**: 683-686.

Tully M & Gilbert HJ (1985) Transformation of *Rhodosporidium toruloides*. *Gene* **36**: 235-240.

Tzfira T, Frankman LR, Vaidya M & Citovsky V (2003) Site-specific integration of *Agrobacterium tumefaciens* T-DNA via double-stranded intermediates. *Plant Physiol* **133**: 1011-1023.

United States Congress (2005) The Energy Policy Act of 2005. Vol. 119 p.^pp. 593.

United States Congress (2007) Energy independence and security act of 2007. Vol. 121 p.^pp. 1492.

van Attikum H & Hooykaas PJ (2003) Genetic requirements for the targeted integration of *Agrobacterium* T-DNA in *Saccharomyces cerevisiae*. *Nucleic Acids Res* **31**: 826-832.

van Kregten M, de Pater S, Romeijn R, van Schendel R, Hooykaas PJ & Tijsterman M (2016) T-DNA integration in plants results from polymerase-theta-mediated DNA repair. *Nat Plants* **2**: 16164.

van Larebeke N, Engler G, Holsters M, Van den Elsacker S, Zaenen J, Schilperoort RA & Schell J (1974) Large plasmid in *Agrobacterium tumefaciens* essential for crown gall-inducing ability. *Nature* **252**: 169-170.

Vergunst AC, Schrammeijer B, den Dulk-Ras A, de Vlaam CMT, Regensburg-Tuink TJG & Hooykaas PJJ (2000) VirB/D4-dependent protein translocation from *Agrobacterium* into plant cells. *Science* **290**: 979-982.

Vogel AM & Das A (1992) Mutational analysis of *Agrobacterium tumefaciens* virD2: tyrosine 29 is essential for endonuclease activity. *J Bacteriol* **174**: 303-308.

Walker JR, Corpina RA & Goldberg J (2001) Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* **412**: 607-614.

Walton FJ, Heitman J & Idnurm A (2006) Conserved elements of the RAM signaling pathway establish cell polarity in the basidiomycete *Cryptococcus neoformans* in a divergent fashion from other fungi. *Mol Biol Cell* **17**: 3768-3780.

Wang QM, Groenewald M, Takashima M, Theelen B, Han PJ, Liu XZ, Boekhout T & Bai FY (2015a) Phylogeny of yeasts and related filamentous fungi within Pucciniomycotina determined from multigene sequence analyses. *Stud Mycol* **81**: 27-53.

Wang QM, Yurkov AM, Goker M, Lumbsch HT, Leavitt SD, Groenewald M, Theelen B, Liu XZ, Boekhout T & Bai FY (2015b) Phylogenetic classification of yeasts and related taxa within Pucciniomycotina. *Stud Mycol* **81**: 149-189.

Wang Y, Lin X, Zhang S, Sun W, Ma S & Zhao ZK (2016) Cloning and evaluation of different constitutive promoters in the oleaginous yeast *Rhodosporidium toruloides*. *Yeast* **33**: 99-106.

Wang ZP, Fu WJ, Xu HM & Chi ZM (2014) Direct conversion of inulin into cell lipid by an inulinase-producing yeast *Rhodosporidium toruloides* 2F5. *Bioresource Technol* **161**: 131-136.

Wiebe MG, Koivuranta K, Penttila M & Ruohonen L (2012) Lipid production in batch and fed-batch cultures of *Rhodosporidium toruloides* from 5 and 6 carbon carbohydrates. *BMC Biotechnol* **12**: 26.

Wiegand I, Hilpert K & Hancock RE (2008) Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat Protoc* **3**: 163-175.

Winkler WC, Nahvi A, Roth A, Collins JA & Breaker RR (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**: 281-286.

Wu S, Hu C, Jin G, Zhao X & Zhao ZK (2010) Phosphate-limitation mediated lipid production by *Rhodosporidium toruloides*. *Bioresour Technol* **101**: 6124-6129.

Wu S, Zhao X, Shen H, Wang Q & Zhao ZK (2011) Microbial lipid production by *Rhodosporidium toruloides* under sulfate-limited conditions. *Bioresour Technol* **102**: 1803-1807.

Xu H, Miao X & Wu Q (2006) High quality biodiesel production from a microalga *Chlorella protothecoides* by heterotrophic growth in fermenters. *J Biotechnol* **126**: 499-507.

Xu JY, Zhao XB, Wang WC, Du W & Liu DH (2012) Microbial conversion of biodiesel byproduct glycerol to triacylglycerols by oleaginous yeast *Rhodosporidium toruloides* and the individual effect of some impurities on lipid production. *Biochem Eng J* **65**: 30-36.

Xu P, Qiao KJ, Ahn WS & Stephanopoulos G (2016) Engineering *Yarrowia lipolytica* as a platform for synthesis of drop-in transportation fuels and oleochemicals. *P Natl Acad Sci USA* **113**: 10848-10853.

Yadav NS, Vanderleyden J, Bennett DR, Barnes WM & Chilton MD (1982) Short direct repeats flank the T-DNA on a nopaline Ti plasmid. *Proc Natl Acad Sci U S A* **79**: 6322-6326.

Yan J, Liu Y, Wang C, Han B & Li S (2015) Assembly of lipase and P450 fatty acid decarboxylase to constitute a novel biosynthetic pathway for production of 1-alkenes from renewable triacylglycerols and oils. *Biotechnol Biofuels* **8**: 34.

Yoshimura T, Okada S & Honda M (2013) Culture of the hydrocarbon producing microalga *Botryococcus braunii* strain Showa: Optimal CO2, salinity, temperature, and irradiance conditions. *Bioresource Technol* **133**: 232-239.

Yusibov VM, Steck TR, Gupta V & Gelvin SB (1994) Association of single-stranded transferred DNA from *Agrobacterium tumefaciens* with tobacco cells. *Proc Natl Acad Sci U S A* **91**: 2994-2998.

Zambryski P, Depicker A, Kruger K & Goodman HM (1982) Tumor induction by *Agrobacterium tumefaciens*: analysis of the boundaries of T-DNA. *J Mol Appl Genet* **1**: 361-370.

Zarnack K, Maurer S, Kaffarnik F, Ladendorf O, Brachmann A, Kamper J & Feldbrugge M (2006) Tetracycline-regulated gene expression in the pathogen *Ustilago maydis*. *Fungal Genet Biol* **43**: 727-738.

Zhang J, Shi L, Chen H, Sun Y, Zhao M, Ren A, Chen M, Wang H & Feng Z (2014) An efficient *Agrobacterium*-mediated transformation method for the edible mushroom *Hypsizygus marmoreus*. *Microbiol Res* **169**: 741-748.

Zhang SY, Skerker JM, Rutter CD, Maurer MJ, Arkin AP & Rao CV (2016) Engineering *Rhodosporidium toruloides* for increased lipid production. *Biotechnol Bioeng* **113**: 1056-1066.

Zheng W, Huang L, Huang J*, et al.* (2013) High genome heterozygosity and endemic genetic recombination in the wheat stripe rust fungus. *Nat Commun* **4**: 2673.

Zhou J, Huang H, Meng K, Shi P, Wang Y, Luo H, Yang P, Bai Y & Yao B (2010) Cloning of a new xylanase gene from *Streptomyces sp*. TN119 using a modified thermal asymmetric interlaced-PCR specific for GC-rich genes and biochemical characterization. *Appl Biochem Biotechnol* **160**: 1277-1292.

Zhou YJ, Buijs NA, Siewers V & Nielsen J (2014) Fatty acid-derived biofuels and chemicals production in *Saccharomyces cerevisiae*. *Front Bioeng Biotechnol* **2**: 32.

Zhou YJ, Buijs NA, Zhu Z, Qin J, Siewers V & Nielsen J (2016) Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. *Nat Commun* **7**: 11709.

Zhu Z, Zhang S, Liu H*, et al.* (2012) A multi-omic map of the lipid-producing yeast *Rhodosporidium toruloides*. *Nat Commun* **3**: 1112.

7 Appendix: synthetic DNA constructs used

7.1.1 Hygromycin phosphotransferase

ATGAAGAAGCCGGAGCTCACCGCCACCTCGGTCGAGAAGTTCCTCATCGAGAAGTTC
GACTCGGTCTCGGACCTCATGCAGCTCTCGGAGGGCGAGGAGTCGCGCGCCTTCTCG
TTCGACGTCGGCGGCCGCGGCTACGTCCTCCGCGTCAACTCGTGCGCCGACGGCTTC
TACAAGGACCGCTACGTCTACCGCCACTTCGCCTCGGCCGCCCTCCCGATCCCGGAG
GTCCTCGACATCGGCGAGTTCTCGGAGTCGCTCACCTACTGCATCTCGCGCCGCGCC
CAGGGCGTCACCCTCCAGGACCTCCCGGAGACCGAGCTCCCGGCCGTCCTCCAGCCG
GTCGCCGAGGCGATGGACGCCATCGCCGCCGCCGACCTCTCGCAGACCTCGGGCTTC
GGCCCGTTCGGCCCGCAGGGCATCGGCCAGTACACCACCTGGCGCGACTTCATCTGC
GCCATCGCCGACCCGCACGTCTACCACTGGCAGACCGTCATGGACGACACCGTCTCG
GCCTCGGTCGCCCAGGCCCTCGACGAGCTCATGCTCTGGGCCGAGGACTGCCCGGAG
GTCCGCCACCTCGTCCACGCCGACTTCGGCTCGAACAACGTCCTCACCGACAACGGC
CGCATCACCGCCGTCATCGACTGGTCGGAGGCCATGTTCGGCGACTCGCAGTACGAG
GTCGCCAACATCTTCTTCTGGCGCCCGTGGCTCGCCTGCATGGAGCAGCAGACCCGC
TACTTCGAGCGCCGCCACCCGGAGCTCGCCGGCTCGCCGCGCCTCCGCGCCTACATG
CTCCGCATCGGCCTCGACCAGCTCTACCAGTCGCTCGTCGACGGCAACTTCGACGAC
GCCGCCTGGGCGCAGGGCCGCTGCGACGCCATCGTCCGCTCGGGCGCCGGCACCGTC
GGCCGCACCCAGATCGCCCGCCGCTCGGCCGCCGTCTGGACCGACGGCTGCGTCGAG
GTCCTCGCCGACTCGGGCAACCGCCGCCCGTCGACCCGCCCGCGCGCCAAGGAGTAG

7.1.2 G418 resistance

ATGGGCAAGGAGAAGACGCACGTCTCGCGCCCGCGCCTCAACTCGAACATGGACGCC
GACCTCTACGGCTACAAGTGGGCCCGCGACAACGTCGGCCAGTCGGGCGCCACGATC
TACCGCCTCTACGGCAAGCCGGACGCCCCGGAGCTCTTCCTCAAGCACGGCAAGGGC
TCGGTCGCCAACGACGTCACGGACGAGATGGTCCGCCTCAACTGGCTCACGGAGTTC
ATGCCGCTCCCGACGATCAAGCACTTCATCCGCACGCCGGACGACGCCTGGCTCCTC
ACGACGGCCATCCCGGGCAAGACGGCCTTCCAGGTCCTCGAGGAGTACCCGGACTCG
GGCGAGAACATCGTCGACGCCCTCGCCGTCTTCCTCCGCCGCCTCCACTCGATCCCG
GTCTGCAACTGCCCGTTCAACTCGGACCGCGTCTTCCGCCTCGCCCAGGCCCAGTCG
CGCATGAACAACGGCCTCGTCGACGCCTCGGACTTCGACGACGAGCGCAACGGCTGG
CCGGTCGAGCAGGTCTGGAAGGAGATGCACAAGCTCCTCCCGTTCTCGCCGGACTCG
GTCGTCACGCACGGCGACTTCTCGCTCGACAACCTCATCTTCGACGAGGGCAAGCTC
ATCGGCTGCATCGACGTCGGCCGCGTCGGCATCGCCGACCGCTACCAGGACCTCGCC
ATCCTCTGGAACTGCCTCGGCGAGTTCTCGCCGTCGCTCCAGAAGCGCCTCTTCCAG
AAGTACGGCATCGACAACCCGGACATGAACAAGCTCCAGTTCCACCTCATGCTCGAC
GAGTTCTTCTAG

7.1.3 Gentamycin resistance

ATGCTCCGCTCGTCGAACGACGTCACGCAGCAGGGCTCGCGCCCGAAGACGAAGCTC
GGCGGCTCGTCGATGGGCATCATCCGCACGTGCCGCCTCGGCCCGGACCAGGTCAAG
TCGATGCGCGCCGCCCTCGACCTCTTCGGCCGCGAGTTCGGCGACGTCGCCACGTAC
TCGCAGCACCAGCCGGACTCGGACTACCTCGGCAACCTCCTCCGCTCGAAGACGTTC

ATCGCCCTCGCCGCCTTCGACCAGGAGGCCGTCGTCGGCGCCCTCGCCGCCTACGTC
CTCCCGCGCTTCGAGCAGCCGCGCTCGGAGATCTACATCTACGACCTCGCCGTCTCG
GGCGAGCACCGCCGCCAGGGCATCGCCACGGCCCTCATCAACCTCCTCAAGCACGAG
GCCAACGCCCTCGGCGCCTACGTCATCTACGTCCAGGCCGACTACGGCGACGACCCG
GCCGTCGCCCTCTACACGAAGCTCGGCATCCGCGAGGAGGTCATGCACTTCGACATC
GACCCGTCGACGGCCACG

7.1.4 EGFP construct[1]

ACCGGCAACAGGATTCAATCTTAAGCCAGACGGACCTTGAGAACCCTCAATCGCTCG
CGGTACTCGTCCGCCCTGCGATCCAGCATCGAAACCGAGTGCAGCGCGTTCAACAAA
TCCGAGTCGTCTCCTCCTGCTCCTTCGCGCTGTTTCGGCGCGGGTGGCGCAGGGACA
GCCGGAGGGAGAGGGGGAGGAGGAGGAGGGGCTGGGGCGACCTTTTTCTTCTTCTTC
CTGTTCTTGCCCTTCTTCTTCGCCGCCTCTGCTTCTCCATCACCCACCGCCCCCACA
TTCGCCGCCGAAGCACCGACCACGGCCCCCCTCATCCGCCTCTGCAACCTCCTCGCC
TCGCTCGCCTCCAAACTCAATCGCGCGACGCACTGCTCCAACTCGGCGATGGCGCTC
ATCAAGCTTGGGAGGGAGGCGGGTGAGAGAGCCGAGTCGGAGAGGATGCCGTCTGCG
GGGATTTGGGAGGGTGAGAGGTGGGTTTGAGGCGGTTGAGAGGAGGAGGACGAGAGG
GGGAGGGCGGAGGAGAGGGCTGTCAAGTCCGAGAGGGAGAGGGGGATCGAGGTCGTG
AAGGATGTCGAGGACGAGGAAGAGGCGTTCCGCTGTTGTTGTTGCTGGACGGCGGAG
AGGACGCCCAGGAAGGCTGCGTCGGCCGCTTGCGAGGGCGAGTCGATCACCTCGTAG
GCGGCGTCGTCGTGCTCGGACGGCGACGACGCGTGGTCCTCCGACTCGACCCAGTCT
GGCCCGTTGAAGAGGCATTCGGCGGCGAAGACGACGGCCGGGTCGTCCGAGGCTGTC
GAGGCAAGTGGCTGTAAGCAAGCGCGGCAGAAGAGGCAGAACGCTGCGACGCACTCC
TAGGATCCCACGGCTGCGCATCCGGATCCTCCCCGCGCTGCTGGCGCCTCGCCCGCT
CCCGCTCGTACTCCGCCATGACCGTCAACCCCTGGCAGAGCAGGCGGTACTGCTCGA
CCAGGGCCCAGTCCTCGCTCGTCTGTGGGAAAGGTTGGACGGGACATCCTGGCGGCA
ATGGAAGCGAGTACAGCCAGTCGAGCGTGAGAGGGGAGGGCCGAAAGTGCTCCGTCG
AGATGCTGGCGAGCGAGACGAGGGGCTGGGTGAGCGACGGCGGGCGGATCCCTGCGC
CTGTGTTGTCCATCCCTGCAGTGCACTCTGTTGCTCGTATCATGTCCCACTCCCTTG
TATCCCTCGAGTCGGTCGACTCTTCCCTGGCGAGTCCAAGCGGAGGAGGTGGTCGTC
GCCTGACCCGCTCGGAGTGCGCCGCTCGACTTGGCCCTGGGAGAACAAGCCTGTGTG
AGTCTGTCTAGCCTGTCAGCGAATGCGCCAGACGAGTGCAAGCGGGTGAGCGAGGTC
GACCCTGCTCGTCACTCGCTCGTCGGGTGCGGCCGCATCGTTGAACTTGCACTTCTC
ACTCGCACTCGCTCTGGTACAGCTACAGTCACTCGCTTACTACTCTGCAGGTTCACA
GCAACTCACCCGTCCAACTCCCACCCTCCCACGTGCAGCCCACCATGGCGGATCCGG
TCTCGAAGGGCGAGGAGCTCTTCACCGGCGTCGTCCCGATCCTCGTCGAGCTCGACG
GCGACGTCAACGGCCACAAGTTCTCGGTCTCGGGCGAGGGCGAGGGCGACGCCACCT
ACGGCAAGCTCACCCTCAAGTTCATCTGCACCACCGGCAAGCTCCCGGTCCCGTGGC
CGACCCTCGTCACCACCCTCACCTACGGCGTCCAGTGCTTCTCGCGCTACCCGGACC
ACATGAAGCAGCACGACTTCTTCAAGTCGGCCATGCCGGAGGGCTACGTCCAGGAGC
GCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTCAAGTTCG
AGGGCGACACCCTCGTCAACCGCATCGAGCTCAAGGGCATCGACTTCAAGGAGGACG
GCAACATCCTCGGCCACAAGCTCGAGTACAACTACAACTCGCACAACGTCTACATCA
TGGCCGACAAGCAGAAGAACGGCATCAAGGTCAACTTCAAGATCCGCCACAACATCG
AGGACGGCTCGGTCCAGCTCGCCGACCACTACCAGCAGAACACCCCGATCGGCGACG
GCCCGGTCCTCCTCCCGGACAACCACTACCTCTCGACCCAGTCGGCCCTCTCGAAGG
ACCCGAACGAGAAGCGCGACCACATGGTCCTCCTCGAGTTCGTCACCGCCGCCGGCA
TCACCCTCGGCATGGACGAGCTCTACAAGTAGTTTCTCCATAATAATGTGTGAGTAG
TTCCCAGATAAGGGAATTAGGGTTCCTATAGGGTTTCGCTCATGTGTTGAGCATATA

AGAAACCCTTAGTATGTATTTGTATTTGTAAAATACTTCTATCAATAAAATTTCTAA
TTCCTAAAACCAAAATCCAGTAGATCTACCATGGTGGACTCCTCTT

[1]Cloning sites in black, *PGK1* promoter in red, EGFP gene in green, CMV35S terminator in blue.

7.1.5 Venus YFP

ATGGACGGCGGCGTCCAGCTCGCCGACCACTACCAGCAGAACACGCCCATCGGAGAC
GGCCCGGTCCTCCTCCCGGACAACCACTACCTCTCGTACCAGTCGGCCCTCTCGAAG
GACCCGAACGAGAAGCGCGACCACATGGTCCTCCTCGAGTTCGTCACGGCCGCCGGC
ATCACGCTCGGCATGGACGAGCTCTACAAGGGCGGCTCGGGCGGCATGGTCTCGAAG
GGCGAGGAGCTCTTCACGGGCGTCGTCCCGATCCTCGTGGAACTCGACGGCGACGTC
AACGGCCACAAGTTCTCGGTCTCGGGCGAGGGCGAGGGCGACGCCACGTACGGCAAG
CTCACGCTCAAGCTCATCTGCACGACGGGCAAGCTCCCGGTCCCGTGGCCGACGCTC
GTCACGACGCTCGGCTACGGCCTCCAGTGCTTCGCCCGCTACCCGGACCACATGAAG
CAGCACGACTTCTTCAAGTCGGCCATGCCGGAGGGCTACGTCCAGGAGCGCACGATC
TTCTTCAAGGACGACGGCAACTACAAGACGCGCGCCGAGGTCAAGTTCGAGGGCGAC
ACGCTCGTCAACCGCATCGAGCTCAAGGGCATCGACTTCAAGGAGGACGGCAACATC
CTCGGCCACAAGCTCGAGTACAACTACAACTCGCACAACGTCTACATCACGGCCGAC
AAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGTAG

7.1.6 Carboxin resistance cassette[2]

CTCACCCGTCCAACTCCCACCCTCCCACGTGCAGCCCACCATGCTCACGCCAGCTGT
CGCACGTGCTGCCCGCGCTGCCCCGAGGGTGCGTCCCCCGTCGTCGTCGCCGTTCTC
CGCCTCCATCCCACCGACTCGCGAGACGCCCCCATCATCGCACTCGCCACACGACTC
CAGGCGGCGCCAGCGACCTCGCCAATGCGTGTACTGACACTGTTTGTTCGACTCGCC
CCGCAGCACGCCCGCACCCTCGCGACCGCCGTCGAGTCGAAGCCGGAGCACCTCAAG
TCGTTCAAGGTCTACCGCTGGGTGAGCCACCTTTGCGGTTGACCGATCTGCAGACTT
GAAGAGAATGCAGGAGTGGCAAAGGGACGAGAGCGAGCGGTGGATAGCTTTGTGGCA
TCTCGAGTTCGGCACTTGGTGGAACCCGCGGCAGGAGATTGGGTGGACTCGAAGCGC
GTGAAGCAGGGCATGGGAGTTCAGGACGACGCGCAACCCCTTTACCGACGCAGGACA
CTCGCGCTGACATCTCCACTGTCCCTCCAACTCCGCGCAGAACCCGGACAAGCCGAC
GGAGAAGCCCTACTTGCAGGAGTACAAGATCGACTTGAACAAGTGCGGCCCGATGGT
CCTCGACGCCATTCTCAAGATCAAGAACGAGCTCGACCCGACCCTTACGTTCCGTCG
CTCGTGCCGTGAGGGAATCTGCGGATCGTGCGCCATGAACATTGACGGTGTGTTCGG
CTTCACACTGATTCGTGCGAGGGGGTTTCGCGCTGACGGGGGATGGTTCGACAGGCG
TCAACACGCTCGCGTGCTTGAAGCGCATCGACAAGGACTCGAAGGATGTCAAGATCT
ACCCCCTCCCGCACAGTGCGTCTCCCGTCCATCTCTCCCTCGCCTCAACCAGCCCGC
TAACTTCGCCTCCCGCCTCGCAGTGTATGTCATCAAGGACCTCGTGCCCGACATGAC
CCAGTGCGTTTCGCTCTGCCCTCGGCGCCACGATCTTCAACTGACTCCGCCCGCTCC
GCAGGTTCTACAAGCAGTACAAGGCGATCCAGCCTTACCTGCAGGCCGACAAGCCCG
CAGACGGTCGCGAGCACCTGCAGTCGAAGGAGGACCGCAAGAAGCTCGACGGCATGT
ACGAGGTGCGTTCCCGCTCCCTCCCTTCTCCCCTTCATTTTCGACTAACACTTACTC
GCTCCGCTCGCCTCGCCTCGCAGTGCATCCTCTGTGCATGCTGCTCGACCTCGTGCC
CCTCCTACTGGTGGAACCAAGACCAGTACCTCGGCCCGGCAGTCCTGATGCAAGCCT
ACCGCTGGGTCGCCGACTCGCGCGACACGCAAAAGGCCGCCCGGCTCGAGAAACTCG
CCAACCCGTTCTCGCTCTACCGCTGCCTCACCATCTTCAACTGCTCCAAGACTTGCC
CTAAATACGTCCCCCACCTTCCTCTCTTTCTCTTACTGTGGTGGTCCGGTCCTGACGA
GTTGCTTTACAGGGTTGAACCCCGCCAAGGCTATCGCCGCACTCAAGCAGGAGATGG
CGTCGGCATGAACTAGTTTTCTCCATAATAATGTGTGAGTAG

[2]Cloning sites in green, *SDH2* gene in black with mutant codon in red.

7.1.7 Carboxin 1kb fragment[3]

GCTGGTGGCAGGATATATTGTGGTGTAAACACGCCTCAACCAGCCCGCTAACTTCGC
CTCCCGCCTCGCAGTGTATGTCATCAAGGACCTCGTGCCCGACATGACCCAGTGCGT
TTCGCTCTGCCCTCGGCGCCACGATCTTCAACTGACTCCGCCCGCTCCGCAGGTTCT
ACAAGCAGTACAAGGCGATCCAGCCTTACCTGCAGGCCGACAAGCCCGCAGACGGTC
GCGAGCACCTGCAGTCGAAGGAGGACCGCAAGAAGCTCGACGGCATGTACGAGGTGC
GTTCCCGCTCCCTCCCTTCTCCCCTTCATTTTCGACTAACACTTACTCGCTCCGCTC
GCCTCGCCTCGCAGTGCATCCTCTGTGCATGCTGCTCGACCTCGTGCCCCTCCTACT
GGTGGAACCAAGACCAGTACCTCGGCCCGGCAGTCCTGATGCAAGCCTACCGCTGGG
TCGCCGACTCGCGCGACACGCAAAAGGCCGCCCGGCTCGAGAAACTCGCCAACCCGT
TCTCGCTCTACCGCTGCCTCACCATCTTCAACTGCTCCAAGACTTGCCCTAAATACG
TCCCCCACCTTCCTCTCTTCTCTTACTGTGGTGGTCCGGTCCTGACGAGTTGCTTTA
CAGGGTTGAACCCCGCCAAGGCTATCGCCGCACTCAAGCAGGAGATGGCGTCGGCAT
GAGCGGACCCGCCGCGCGGCAGAAGGAGGCGGACATTTTCAGAGGGGACGCCAGCAG
TAGCGATCGTGCTTGTAACTTCTCGTTTTTCCGTGCGTCGTCCAGCTTGGCCCTTCG
GGAGCGGATAGTTTGCGTCCCTTGTGCGTGGGAGATGGCGGAGTGGGGCGAAATGGA
CGAAGCGAGCGCGGCGGCGAGAACAGCAAGGTAGCGGGATCCTCTCGCCTACGTTGG
TTCGTCATGGTCGATCGAATCACCTCACTTCAGCTCGCACCACTACAGCAGCAATGT
GTAAGACTGAGCAGAGAGTGGGAAAAGGCGCCTCGGAGCGAAGAGCGACGCCGGAAG
AGAGGGGTAAACCTAAGAGAAAAGAGCGTTTAGATTT

[3]Cloning sites in green, *SDH2* gene in black with mutant codon in red, downstream sequence in orange.

7.1.8 OleT[4]

CCACCCTCCCACGTGCAGCCCACCATGGCCACCCTCAAGCGCGACAAGGGCCTCGAC
AACACCCTCAAGGTCCTCAAGCAGGGCTACCTCTACACCACCAACCAGCGCAACCGC
CTCAACACCTCGGTCTTCCAGACCAAGGCCCTCGGCGGCAAGCCGTTCGTCGTCGTC
ACCGGCAAGGAGGGCGCCGAGATGTTCTACAACAACGACGTCGTCCAGCGCGAGGGC
ATGCTCCCGAAGCGCATCGTCAACACCCTCTTCGGCAAGGGCGCCATCCACACCGTC
GACGGCAAGAAGCACGTCGACCGCAAGGCCCTCTTCATGTCGCTCATGACCGAGGGC
AACCTCAACTACGTCCGCGAGCTCACCCGCACCCTCTGGCACGCCAACACCCAGCGC
ATGGAGTCGATGGACGAGGTCAACATCTACCGCGAGTCGATCGTCCTCCTCACCAAG
GTCGGCACCCGCTGGGCCGGCGTCCAGGCCCCGCCGGAGGACATCGAGCGCATCGCC
ACCGACATGGACATCATGATCGACTCGTTCCGCGCCCTCGGCGGCGCCTTCAAGGGC
TACAAGGCCTCGAAGGAGGCCCGCCGCCGCGTCGAGGACTGGCTCGAGGAGCAGATC
ATCGAGACCCGCAAGGGCAACATCCACCCGCCGGAGGGCACCGCCCTCTACGAGTTC
GCCCACTGGGAGGACTACCTCGGCAACCCGATGGACTCGCGCACCTGCGCCATCGAC
CTCATGAACACCTTCCGCCCGCTCATCGCCATCAACCGCTTCGTCTCGTTCGGCCTC
CACGCCATGAACGAGAACCCGATCACCCGCGAGAAGATCAAGTCGGAGCCGGACTAC
GCCTACAAGTTCGCCCAGGAGGTCCGCCGCTACTACCCGTTCGTCCCGTTCCTCCCG
GGCAAGGCCAAGGTCGACATCGACTTCCAGGGCGTCACCATCCCGGCCGGCGTCGGC
CTCGCCCTCGACGTCTACGGCACCACCCACGACGAGTCGCTCTGGGACGACCCGAAC
GAGTTCCGCCCGGAGCGCTTCGAGACCTGGGACGGCTCGCCGTTCGACCTCATCCCG
CAGGGCGGCGGCGACTACTGGACCAACCACCGCTGCGCCGGCGAGTGGATCACCGTC
ATCATCATGGAGGAGACCATGAAGTACTTCGCCGAGAAGATCACCTACGACGTCCCG
GAGCAGGACCTCGAGGTCGACCTCAACTCGATCCCGGGCTACGTCAAGTCGGGCTTC

GTCATCAAGAACGTCCGCGAGGTCGTCGACCGCACCTAGTTTCTCCATAATAATGTG
TGAGTAG

[4]Cloning sites in red, *OleT* gene in black.

7.1.9 Lip2[5]

CCACCCTCCCACGTGCAGCCCACCATGCGCTCGTCGCTCGTCCTCTTCTTCGTCTCG
GCCTGGACCGCCCTCGCCTCGCCGATCCGCCGCGAGGTCTCGCAGGACCTCTTCAAC
CAGTTCAACCTCTTCGCCCAGTACTCGGCCGCCGCCTACTGCGGCAAGAACAACGAC
GCCCCGGCCGGCACCAACATCACCTGCACCGGCAACGCCTGCCCGGAGGTCGAGAAG
GCCGACGCCACCTTCCTCTACTCGTTCGAGGACTCGGGCGTCGGCGACGTCACCGGC
TTCCTCGCCCTCGACAACACCAACAAGCTCATCGTCCTCTCGTTCCGCGGCTCGCGC
TCGATCGAGAACTGGATCGGCAACCTCAACTTCGACCTCAAGGAGATCAACGACATC
TGCTCGGGCTGCCGCGGCCACGACGGCTTCACCTCGTCGTGGCGCTCGGTCGCCGAC
ACCCTCCGCCAGAAGGTCGAGGACGCCGTCCGCGAGCACCCGGACTACCGCGTCGTC
TTCACCGGCCACTCGCTCGGCGGCGCCCTCGCCACCGTCGCCGGCGCCGACCTCCGC
GGCAACGGCTACGACATCGACGTCTTCTCGTACGGCGCCCCGCGCGTCGGCAACCGC
GCCTTCGCCGAGTTCCTCACCGTCCAGACCGGCGGCACCCTCTACCGCATCACCCAC
ACCAACGACATCGTCCCGCGCCTCCCGCCGCGCGAGTTCGGCTACTCGCACTCGTCG
CCGGAGTACTGGATCAAGTCGGGCACCCTCGTCCCGGTCACCCGCAACGACATCGTC
AAGATCGAGGGCATCGACGCCACCGGCGGCAACAACCAGCCGAACATCCCGGACATC
CCGGCCCACCTCTGGTACTTCGGCCTCATCGGCACCTGCCTCTAGTTTCTCCATAAT
AATGTGTGAGTAG

[5]Cloning sites in red, *Lip2* gene in black.

7.1.10 Synthetic gene cluster containing codon optimised *Synechococcus elongatus* aldehyde decarbonylase, *E. coli* ferredoxin and ferredoxin reductase under the regulation of the *R. toruloides* CBS 14 *TUB1*, *THI5* and *THI4* constitutive promoters respectively[6]

ATCATCGCAAGACCGGCAACAGGATTCAATGAATTCTAAGCGCCCAGGCGTGTTAGG
TTGATGGACGGTTCGCCAAAAGAGGGAGACGAGAATCATATCCTACCCCCTAGACTT
CTTCTGACTCTTCTTCTCTGATTCGGTTTGCCCTCGACAAACGAACAACACACCTTG
TTGCAAAGCACCCTTCTCACCCCATTTCAGCGCGTCGCTCAGCGTCATCTGCTCCAT
TCGTCGCCGCTTCTGCGTTCCTCTGCCCGTTCGCTCCGCTCTTTCTCCTTATCCCAC
GCGCTCACTCCTCTCAGCGCAATCTAGCATGCAGCGTCAACGCTCCTCAGTCGAGGA
CGCACATGCTGCCCCCCGAATGAACGAATCAGAATGATGTTGCGCTTCGTCGCCTCT
CTCGCCGACTCGTCAGTAGCATAGGTCTCGCACGACTCCCCGAGATGGATGAGGCGC
AACGCCCGGTAGTTTTGGTTGCAAGGACAGTCACAGGTCAACCGCCTCGGTAAGCGA
GAGCGGCGGGTGGAGATCTACAGAGACACGCGCGCGAGAGAGAGGAACGAAGCGAGA
ACCGAAGCGAGAGACTGCTCACGGAAGGTCGACGAGCTTGAATGGCGGGCTGGCGGG
TCGCTCTGGTACGCGACCAGGACGAACTGGGCCAGGCGAGGGAGAGGATGTCGAGCG
AGGTCGCGATCGCCCCGATGCAAGCGTTTCAACGCTGGCCTCGCTGGCGCCGCTTCG
CTGGCGAGCGTGACCAGGCCGGCGCAGGGCGAGCTTCTCGATGACCTCCTTCGCTGC
GCCTGCGAGTTTCTCGGCAGCGGTAGTGCTTCCGCTGCTTGCCGTCCGGCTCCTCGA
GGCGGAGTGGGAGCGGCGCGGCCTGCGCTGCTGGTCTTCGAGTGAAGGCCGACGACG
ACTGCGCGTCTCGGCGTCGTCCGACGCGGAGCGTCTGTCGCCATCCGCATCGTGCCG
CTTGGCTGTCGAAGACATGCGACGATGGTATCGCTTCGTCGCCGCGGCGAGGGGAGA
GGCGGGAGGTGAAGGCGACGAGTTTGCCGAGGTGAAGATACGCTGAAATGGACGAGG

TGGAGACTTGATGATGAGCGTGTCGGCGGAGAAACGCGACGTGAGGGTGCGGTGCTG
AGGCTCGGGCGGTGGAGGAGGAGGAGGAGGAGGATACGCTGGCGGGAAGCGCAAGGC
CGCCTTGAGGGTGCTCATCTGTGGTGATGCTCAGTCAGCGTCGTCTTCTCCAATATG
GGTGGAAGTGAACTCACACTTCGCTGAGAGGGCCGTTGCGCTGGTGACGAGGAGCAG
GAAAAAGGAGCAGCAGAACACGCGGATGGATGGAAGCAAGTGCATCCTCGAGGCGAT
CGGCGCGAGGGAACGTCGCATAGGCTGCCGTCAACAAGGCGTCGCACGGACAAAGGA
GACAAGGCGGGTAGCAGAGCTCGGTCGCTGGACAGGAGCCTCTGCCTGAGCGTCAGG
AGGCCCGTGAGCGAGGACAAGGTGCGGGAGTGATTAGAGCGGATGACGATGGCTGTA
CGGTACAGGCAACCTCCGCAAAATCTCGTCAGTCTGGGTCGCACGTTGCAGCGCATT
GCCTCCGGTAGCGAGCAAGAGTCGACGGAGACGGATATCGAGTGAGAGGGGACGAGG
CGGTGAGGGGTTTGGCGGCGCTGTGATAGCTGCGACGGGTGTTGTAAGCCTCGCAGA
TTGCCTCTCCGTCGGTGATAGCAGCTCATCAGCCCCAGCGTGGGTCAGCGTGAGACG
CGAAAGGGTTTGGATCGGGTCGCTCGCTCCTTGTCCTCACCCGCACCGCTCAGCAGT
CCCGTCGTAGCTTCACGGACCAGCCAGAGCCCCTCAGACCTTTCCACCTACACTGTG
AGTCGCCTTGAACATCGCAAACAGCCCCCGCGCGACGGTGTCGTCGATGAAACGATG
AAACGCGCTCTCGTCCCACCCCTCCCTCGCTCAACGTCGCTGACCCACTCCCACACA
CACAGTCCCGCACCCCGACAACCAGTCTCCCCTCTTCCTCGAGTCACCATGCCGCAG
CTCGAGGCCTCGCTCGAGCTCGACTTCCAGTCGGAGTCGTACAAGGACGCCTACTCG
CGCATCAACGCCATCGTCATCGAGGGCGAGCAGGAGGCCTTCGACAACTACAACCGC
CTCGCCGAGATGCTCCCGGACCAGCGCGACGAGCTCCACAAGCTCGCCAAGATGGAG
CAGCGCCACATGAAGGGCTTCATGGCCTGCGGCAAGAACCTCTCGGTCACCCCGGAC
ATGGGCTTCGCCCAGAAGTTCTTCGAGCGCCTCCACGAGAACTTCAAGGCCGCCGCC
GCCGAGGGCAAGGTCGTCACGTGCCTCCTCATCCAGTCGCTCATCATCGAGTGCTTC
GCCATCGCCGCCTACAACATCTACATCCCGGTCGCCGACGCCTTCGCCCGCAAGATC
ACCGAGGGCGTCGTCCGCGACGAGTACCTCCACCGCAACTTCGGCGAGGAGTGGCTC
AAGGCCAACTTCGACGCCTCGAAGGCCGAGCTCGAGGAGGCCAACCGCCAGAACCTC
CCGCTCGTCTGGCTCATGCTCAACGAGGTCGCCGACGACGCCCGCGAGCTCGGCATG
GAGCGCGAGTCGCTCGTCGAGGACTTCATGATCGCCTACGGCGAGGCCCTCGAGAAC
ATCGGCTTCACCACCCGCGAGATCATGCGCATGTCGGCCTACGGCCTCGCCGCCGTC
TAGAGTCCCGTAGTGTGCGGACGGGACTTTGGAGTTTTGTTCACCCTCTCTTCCCTT
CTTCTGTTCCCCTCTTCGCTTTTCCTTGCGGTAATGCCAGTTTGATGCGCCTCTCTA
CTGTGTGTGCTCTACTGAGCCTGAGATGCAGTAAGCGGCAAGGCGTCCGCAGGGCAT
CCGCGAGGGAGAGGCGACACAATGGGGCGCCGCGGGCTCACTCTGCGGTTTCAGCGT
GCGCCTCACAGTTCTGGACAATCCAGAGCTTGCGTCCGTCTCTCGGTCGTCTCGCTG
TTTCGCCCCGCTCGTCCTAACCCTCGTCCGCTCGCAATAGCTTCATGGTCCAGGGCG
GCGACTTTACGATGCGCAACGGCAAAGGCGGCGAGTCGATCTACGGCCAGACGTTCG
AGGACGAGGACCTGAGGCGCGAGATTGATTCGGAAGGGTTGTTGTGCATGGCGAACA
AGGGACGCAAGTGCGTTCGTTCTTTTCGTGGCGATGCCTCTCATCTTCACCAAGATC
GTTGCGGTCGTCTCGAGCGGCTGACGAGGCCGTAAGTCCTACTCGGTTGTGCTGACA
CCGTTTCTCACCTCGCAGCACCAACTCCTCTCAGTTCTTCGTCACCCTGCGCCCTTG
CCCTCACCTGAACGGCAAACACGTTGTCTTCGGCAAGGTGGTCAAGGGTGCGTGTGG
ATTGTCCAAACAAGGCCTGCCCTGCTGATGGAAGTCTCGCACGCAGGCTACGAAATC
ATCGTCGCGATGAGCAAGATGCCGGTCGACGCCAAGGACCACCCAACACAGCTCATT
ACGATCTCGCATTGCGGCGAGCTCGAGCGCCGCGTCGCGGCGAAACCCAAAACCCCT
CCTCCCGCCTCGCCATCCGCCACTTCCCGCTCGCGGTCTCGCTCTCGCTCCGTCTCT
CGCTCGCCTTCGCCCGACCGCAATTGGTCCCGTTCATCCCGCCACCGCTCTCGCCGC
CACGACACCGATGAGTCTGGCTCAGAAGACGACGACCGTCGCCGCCGTCATCGCTCG
TCCTCGCACAAACACCGCTCTTCTCGACGACACAAGCACCGCTCGTCTCGTCGCTCA
GCGTCTCCTGCGCGGAAGGAGGACTCGCCCGAGCTCTCAGCGGAACAGATTGCGGCG
CTAGAGGCGCAAAATAGGGAGGAAGAAACTGCGGCGAAAAGGAGACGTGAGGAGGAG
GAGCGCGAGGCAGAGTTGCGTAGGCGCGAGGCGCAGGAAGCGAGGGAGCGCTACGAG

CGGGAGCAGATCGCGAAGGGCGGGATCATCTACAAGGGAAGGGGGCGGATGAAGGCG
GATAGCGGTAGGGGTGGAATGAGGGGCTGGTAATTCGACGGTGTCTCGACGCCTTCG
ACTTCTTTTCGACTTCGCTTCCCCCTTTGCTTGTCCGGCCCGTCCAGCTCCGCCCGA
CAACGACCGGCTGGTGGAGGATTGGCAGGCGGTATCGTCGGCATCGGTTGGCCTTCT
CAACCTTGCTCGCTCTTCTTGTGCTGCTCTCACGACGTCGCTTGCTTTCGCCTTGAC
GTTGGGCGAGCACGTTTGACTGCGAGGAATCGATACGGGGTCGATGTATGCGGTGCT
CGTGCCGAAAAACGCTCCTACTCCCGCCGAAGCGCAACGGCCCACGACACCTCGAAA
ACCTCTCTTTCTCTTCTTCTTCTTCCTTCCCTAGCACCACAGCGTCAATACAACGTC
GATCCCAACCCACCCGCACTCCTACGACACCTACAGCACTAGCAGAATGGCCGACTG
GGTCACCGGCAAGGTCACCAAGGTCCAGAACTGGACCGACGCCCTCTTCTCGCTCAC
CGTCCACGCCCCGGTCCTCCCGTTCACCGCCGGCCAGTTCACCAAGCTCGGCCTCGA
GATCGACGGCGAGCGCGTCCAGCGCGCCTACTCCTACGTCAACTCGCCGGACAACCC
GGACCTCGAGTTCTACCTCGTCACCGTCCCGGACGGCAAGCTCTCGCCGCGCCTCGC
CGCCCTCAAGCCGGGCGACGAGGTCCAGGTCGTCTCGGAGGCCGCCGGCTTCTTCGT
CCTCGACGAGGTCCCGCACTGCGAGACCCTCTGGATGCTCGCCACCGGCACCGCCAT
CGGCCCGTACCTCTCGATCCTCCAGCTCGGCAAGGACCTCGACCGCTTCAAGAACCT
CGTCCTCGTCCACGCCGCCCGCTACGCCGCCGACCTCTCGTACCTCCCGCTCATGCA
GGAGCTCGAGAAGCGCTACGAGGGCAAGCTCCGCATCCAGACCGTCGTCTCGCGCGA
GACCGCCGCCGGCTCGCTCACCGGCCGCATCCCGGCCCTCATCGAGTCGGGCGAGCT
CGAGTCGACCATCGGCCTCCCGATGAACAAGGAGACCTCGCACGTCATGCTCTGCGG
CAACCCGCAGATGGTCCGCGACACCCAGCAGCTCCTCAAGGAGACCCGCCAGATGAC
CAAGCACCTCCGCCGCCGCCCGGGCCACATGACCGCCGAGCACTACTGGTAGTAAGC
GGTTTTAGACTCCCGCCCCTCTCGCCTAGCCTCTTCGCTTCTTCCACTGCGCAGTTT
GAGTGAATGCACAAAATTCAGTTCACAGCCTGTGCGTTGACGATGCGGGAGAGATGA
CGGAGTCGAGAGATAGACGTCTGCCTCTCGCTTCCCCTCGTCCGCGCTAGCCTAAGT
CTGCCTCCGCCTTCGTTCCCGTTGCTCATCGGCGCAAGCCAGCATCCTCTCGAGTGC
GACGAGCATACTACGAACTGATTCGGACGTCCTTTTGCACCGAAAACCACTCGAGCG
CGACGAAGAGAGAGACAGGAAGCCAAGCTAACTACAACTTTGCGGTACCCGTAGCCG
GGTCAGCGAGTCAGCAGAGCAAGAAGAACCGCGACTTCAGTTCGTCCCGCCCCTCTC
TCTCTCGCTGACACGCCAGGAACTGATGCGGTGGTCGCAGTCTTTGAGCGCACTCGC
TCCGGAGGATTCGTCCAGAACGATGTGCTTGTTCAGTTCATGATCCCTGGCTTGCCG
TTCGGCGGTACGGGCGCGGCGGGCTACGGAAACTACCACGGTAGGCGGTGCGTCCTC
CCTCTCTTACCTTTCTTCCGTCTCGGCCGGTGACTGACTCGCTTCCCCGGGCGAGCA
GCACCTTCGACACGTTCTCCCACGAGCGCGCGTCGGCCAACGTCCCGACCTGGATGG
ACATGATCATGGCGTCGCGGTACCCGCCCTACACCCGTTCGTCGCATTCCTCCGCTC
CTTCCTCGCGCTCGCTGACACCCTCGCTCCACTCCACGCACGCAGAGAAGAAGCTCA
AGATGCTCCTGTTCGCGACCAAGGCGGTGATCAAGAAGCCCAGCAAGTTTGGCTCGA
TCTCGCGCCTGCTCAAGGTGATTGCCGCGATGGTCGCTCTCTTGGCTGTCAGGGCAA
GGCTCTGACTGACCCGTCGTCGTCCTACTCCATCCCTTCTCCCCTCCCCTTTCTCTC
GGGTCTTGGGTCGCGTGTGCGTCGGGCAGAAGTTGACGGGACAGGCGTGAGGCGGGA
ACAAGAGTCGCCCATAGCTCTTCAGATCTATTCACTCCACTCAATCAAGTGTTGTCT
GTACCCTCCCCATCTCCCTCCCAATCAGTTGTCTGTACCGGCACCGCCGCGAAAACC
CCAAAGTCTGTACCGTCCATTCGCAACGAGTTTGTCCCGCAACGTCGTTCCGCGGTC
TCGTCCGTCTTTCTGCGGTTGTGAGCGAGGATCAAACGCTGTGAGAGCCAGCGGGTG
CACTGCGAAGGGGACGTGGAGGTGGAGCAGGAGCGATAGGGTTGCGACTTGAGCGTG
GAGGTCGTGCCGAAGCGGAGAGGACGGGTGCGTCTCGCAGCGAGGGTGTGGGGAGCT
GTACGCCCGCACGGAAACCACGGTGACGCTGCACGGTCGAAAGCAGAGGGCTGAGAA
GTGAGAGAGAGGCGAAGGGGCGAAGCGATGAGGAGCTTCTCGACGAGCGTGCCGGAC
GAAGAGCGCTTCGCAGGGCTAGCTGGCGGTGCGTGCGGCATAACTGAGTAGTACACA
GGTCCTGATAGCAGCGGTGGCCCCCCAAAGTGCTTGTGAGGCGAGCGCGAGCGACCG
CAACCCGCCGGCGCACAGTCTTCCCCTGGCTGACCTTCCCAAGCTCGCTTCTTCGGC

AGGCGAGGGACCCCAACTTGGCGCCTGGGGCGGTGCTCGATGGCGCGATTCCACGAG
TCAAACGAGAGGTGAAGTGTATGCACTTTGCCGATAGGAGCCTGGCGCGACTCGTGG
GACTCGTTCGAGGGTTGCTCCGCTCCTTCTTCTTCCCTCACCTCAGTACCGCTCCTA
CGCACGCTCATCGACGTTTAAGAATCAACATGCCGAAGATCGTCATCCTCCCGCACC
AGGACCTCTGCCCGGACGGCGCCGTCCTCGAGGCCAACTCGGGCGAGACCATCCTCG
ACGCCGCCCTCCGCAACGGCATCGAGATCGAGCACGCCTGCGAGAAGTCGTGCGCCT
GCACCACGTGCCACTGCATCGTCCGCGAGGGCTTCGACTCGCTCCCGGAGTCGTCGG
AGCAGGAGGACGACATGCTCGACAAGGCCTGGGGCCTCGAGCCGGAGTCGCGCCTCT
CGTGCCAGGCCCGCGTCACCGACGAGGACCTCGTCGTCGAGATCCCGCGCTACACCA
TCAACCACGCCCGCGAGCACTAGTTTCTCCATAATAATGTGTGAGTAGTTCCCA

[6]*CYC1* in dark blue, *TUB1* promoter in light green, *Synechococcus elongates* Aldehyde decarbonylase in brown, *TUB1* terminator in light blue, *NMT1* promoter in dark purple, *E. coli* Ferredoxin reductase in red, *GPD1* terminator in dark green, *NMT2* Promoter in light purple, *E. coli*, Ferredoxin in orange.

7.1.11 *Synechococcus sp.* fatty acyl-ACP reductase[7]

GAATTCCAACTCCCACCCTCCCACGTGCAGCCCACCATGTTCGGCCTCATCGGCCAC
CTCACCTCGCTCGAGCAGGCCCGCGACGTCTCGCGCCGCATGGGCTACGACGAGTAC
GCCGACCAGGGCCTCGAGTTCTGGTCGTCGGCCCCGCCGCAGATCGTCGACGAGATC
ACCGTCACCTCGGCCACCGGCAAGGTCATCCACGGCCGCTACATCGAGTCGTGCTTC
CTCCCGGAGATGCTCGCCGCCCGCCGCTTCAAGACCGCCACCCGCAAGGTCCTCAAC
GCCATGTCGCACGCCCAGAAGCACGGCATCGACATCTCGGCCCTCGGCGGCTTCACC
TCGATCATCTTCGAGAACTTCGACCTCGCCTCGCTCCGCCAGGTCCGCGACACCACC
CTCGAGTTCGAGCGCTTCACCACCGGCAACACCCACACCGCCTACGTCATCTGCCGC
CAGGTCGAGGCCGCCGCCAAGACCCTCGGCATCGACATCACCCAGGCCACCGTCGCC
GTCGTCGGCGCCACCGGCGACATCGGCTCGGCCGTCTGCCGCTGGCTCGACCTCAAG
CTCGGCGTCGGCGACCTCATCCTCACCGCCCGCAACCAGGAGCGCCTCGACAACCTC
CAGGCCGAGCTCGGCCGCGGCAAGATCCTCCCGCTCGAGGCCGCCCTCCCGGAGGCC
GACTTCATCGTCTGGGTCGCCTCGATGCCGCAGGGCGTCGTCATCGACCCGGCCACC
CTCAAGCAGCCTTGCGTTCTCATCGACGGCGGCTACCCGAAGAACCTCGGCTCGAAG
GTCCAGGGCGAGGGCATCTACGTCCTCAACGGCGGCGTCGTCGAGCACTGCTTCGAC
ATCGACTGGCAGATCATGTCGGCCGCCGAGATGGCCCGCCCGGAGCGCCAGATGTTC
GCCTGCTTCGCCGAGGCCATGCTCCTCGAGTTCGAGGGCTGGCACACCAACTTCTCG
TGGGGCCGCAACCAGATCACCATCGAGAAGATGGAGGCCATCGGCGAGGCCTCGGTC
CGCCACGGCTTCCAGCCGCTCGCCCTCGCCATCTAGTAAGCGCCCAGGCGTGTTAGG
TTGATGGACGAATTCTTTCTCCATAATAATGTGTGAGTAGTTCCC

[7]Cloning sites in blue, coding sequence in black.

7.1.12 *Acinetobacter sp.* fatty acyl-CoA reductase[8]

GAATTCCAACTCCCACCCTCCCACGTGCAGCCCACCATGAACAAGAAGCTCGAGGCC
CTCTTCCGCGAGAACGTCAAGGGCAAGGTCGCCCTCATCACCGGCGCCTCGTCGGGC
ATCGGCCTCACCATCGCCAAGCGCATCGCCGCCGCCGGCGCCCACGTCCTCCTCGTC
GCCCGCACCCAGGAGACCCTCGAGGAGGTCAAGGCCGCCATCGAGCAGCAGGGCGGC
CAGGCCTCGATCTTCCCGTGCGACCTCACCGACATGAACGCCATCGACCAGCTCTCG
CAGCAGATCATGGCCTCGGTCGACCACGTCGACTTCCTCATCAACAACGCCGGCCGC
TCGATCCGCCGCGCCGTCCACGAGTCGTTCGACCGCTTCCACGACTTCGAGCGCACC
ATGCAGCTCAACTACTTCGGCGCCGTCCGCCTCGTCCTCAACCTCCTCCCGCACATG
ATCAAGCGCAAGAACGGCCAGATCATCAACATCTCGTCGATCGGCGTCCTCGCCAAC
GCCACCCGCTTCTCGGCCTACGTCGCCTCGAAGGCCGCCCTCGACGCCTTCTCGCGC

TGCCTCTCGGCCGAGGTCCTCAAGCACAAGATCTCGATCACCTCGATCTACATGCCG
CTCGTCCGCACCCCGATGATCGCCCCGACCAAGATCTACAAATACGTCCCGACCCTC
TCGCCGGAGGAGGCCGCCGACCTCATCGTCTACGCCATCGTCAAGCGCCCGAAGCGC
ATCGCCACCCACCTCGGCCGCCTCGCCTCGATCACCTACGCCATCGCCCCGGACATC
AACAACATCCTCATGTCGATCGGCTTCAACCTCTTCCCGTCGTCGACCGCCGCCCTC
GGCGAGCAGGAGAAGCTCAACCTCCTCCAGCGCGCCTACGCCCGCCTCTTCCCGGGC
GAGCACTGGTAGTAAGCGCCCAGGCGTGTTAGGTTGATGGACGAATTCTTTCTCCAT
AATAATGTGTGAGTAGTTCCC

[8]Cloning sites in blue, coding sequence in black.