Development of a genetic and molecular toolkit for the oleaginous red yeast *Rhodotorula toruloides*

Submitted by George Alexander Kirke to the University of Exeter as a thesis for the degree of Master of Science by Research (Biological Sciences), April 2019.

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

(Signature) …………………………………………………………………………………………………………………
Abstract

Rhodotorula toruloides is an oleaginous yeast with potential use as a biotechnological chassis for both production of industrially and pharmaceutically relevant compounds, and as a drop-in biofuel producer on low-cost substrate. Cells can accumulate lipid droplets to over 70 % weight/weight under certain growth conditions. We here summarise the currently-available genetic and molecular toolkit for the yeast and suggest further avenues for research to enable full utilisation of this yeast.

To aid in these objectives, we have constructed lipid droplet-associated GFP-tagged protein Ldp1-GFP and demonstrated how this can be used to quantify individual cell lipid quantity using confocal microscopy. Calnexin-GFP and GFP-Atg8 have also been constructed for live-cell monitoring of the intracellular machinery in lipid droplet synthesis as part of the developing R. toruloides molecular toolkit.

Furthermore, the induction profiles of selected heat shock protein promoters have been characterised through a GFP-reporter system. Currently, the only reported inducible promoters in R. toruloides are nutrient-dependent NAR1, ICL1, MET16, CTR3, DAO1, THI4, THI5 and CTR31, and tightly controlling these is not possible in potential use as a biofuel producer using low-cost substrate. Of the promoters highlighted herein (ENO2, TDH3, PGK1, TPI1, SSB1, ACT1 and TDH3), ENO2 is identified as a qualitatively putative heat-shock inducible promoter, further analysis of which may allow the nutrient-dependency limitation to be overcome through exploitation of the native environmental stress response.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>2</td>
</tr>
<tr>
<td>Contents</td>
<td>3</td>
</tr>
<tr>
<td>Abbreviations</td>
<td>5</td>
</tr>
<tr>
<td>List of Figures</td>
<td>6</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>8</td>
</tr>
<tr>
<td>1.1. Molecular process and organelle targeting</td>
<td>12</td>
</tr>
<tr>
<td>1.1.1. Lipid biogenesis</td>
<td>13</td>
</tr>
<tr>
<td>1.1.2. Lipid droplets</td>
<td>21</td>
</tr>
<tr>
<td>1.2. Heat shock promoters</td>
<td>26</td>
</tr>
<tr>
<td>1.3. Project aims</td>
<td>30</td>
</tr>
<tr>
<td>2. Materials and Methods</td>
<td>31</td>
</tr>
<tr>
<td>2.1. Strains and growth conditions</td>
<td>31</td>
</tr>
<tr>
<td>2.1.1. <em>Rhodotorula toruloides</em></td>
<td>31</td>
</tr>
<tr>
<td>2.1.2. Bacteria</td>
<td>31</td>
</tr>
<tr>
<td>2.1.3. <em>Saccharomyces cerevisiae</em></td>
<td>31</td>
</tr>
<tr>
<td>2.2. Bioinformatics</td>
<td>32</td>
</tr>
<tr>
<td>2.2.1. <em>R. toruloides</em> homologous gene identificiation</td>
<td>32</td>
</tr>
<tr>
<td>2.2.2. Codon Optimisation</td>
<td>32</td>
</tr>
<tr>
<td>2.3. Plasmid construction</td>
<td>32</td>
</tr>
<tr>
<td>2.3.1. <em>R. toruloides</em> genomic DNA extraction</td>
<td>32</td>
</tr>
<tr>
<td>2.3.2. Amplification of DNA</td>
<td>32</td>
</tr>
<tr>
<td>2.3.3. Restriction digestion</td>
<td>32</td>
</tr>
<tr>
<td>2.3.4. Plasmid recovery from micro-organisms</td>
<td>33</td>
</tr>
<tr>
<td>2.3.5. DNA gel electrophoresis</td>
<td>33</td>
</tr>
<tr>
<td>2.3.6. Bacterial transformation</td>
<td>33</td>
</tr>
<tr>
<td>2.3.7. <em>S. cerevisiae</em> transformation</td>
<td>34</td>
</tr>
</tbody>
</table>
2.4. Transformation of *R. toruloides* ................................................................. 37
2.5. CARS/SRS ........................................................................................................ 37
2.6. Confocal microscopy .......................................................................................... 37
2.7. Quantification of lipid droplets in confocal fluorescence images ........ 38
2.8. Quantification of yeast growth rate and lipid droplet responses to different media .................................................................................................................................. 39
2.9. Measuring single-cell HSPP induction ............................................................. 40
2.10. Measuring HSPP induction kinetics ............................................................... 40

3. Results & Discussion: Organelle visualisation in *R. toruloides* ......... 41
3.1. Construction of GFP-tagged protein expression plasmids .................. 41
3.2. Visualisation of lipid droplets in *R. toruloides* ........................................... 47
3.3. Development of a precise and a rapid method for quantification of lipid biogenesis ................................................................. 50
3.4. Quantification of lipid biogenesis under glucose starvation conditions 51
3.5. Endoplasmic reticulum visualisation in *R. toruloides* ............................. 54
3.6. Autophagy visualisation and induction in *R. toruloides* ........................ 55

4. Results & Discussion: Heat Shock Protein Promoter Induction ........ 58
4.1. Promoter design and synthesis ........................................................................ 58
4.2. Construction of HSPP reporter plasmids ...................................................... 63
4.3. Heat shock promoter induction in *R. toruloides* ....................................... 67
4.4. Heat Shock promoter kinetics in *R. toruloides* .......................................... 69

5. Conclusion ............................................................................................................ 74

6. Acknowledgements ............................................................................................. 75

7. Bibliography ......................................................................................................... 76
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGPAT</td>
<td>Acylglycerol-3-phosphate acyltransferase</td>
</tr>
<tr>
<td>ASTM International</td>
<td>(formerly American Society for Testing and Materials) International</td>
</tr>
<tr>
<td>ATMT</td>
<td>Agrobacterium tumefaciens-mediated transformation</td>
</tr>
<tr>
<td>CARS</td>
<td>Coherent anti-Stokes Raman spectroscopy</td>
</tr>
<tr>
<td>CSM</td>
<td>Complete Supplement Mix</td>
</tr>
<tr>
<td>DAG</td>
<td>Diacylglycerol</td>
</tr>
<tr>
<td>DGAT</td>
<td>Diacylglycerol acyltransferase</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>FA</td>
<td>Fatty acid</td>
</tr>
<tr>
<td>FAMEs</td>
<td>Fatty acid methyl esters</td>
</tr>
<tr>
<td>FAS</td>
<td>Fatty acid synthase</td>
</tr>
<tr>
<td>FFA</td>
<td>Free fatty acid</td>
</tr>
<tr>
<td>GHG</td>
<td>Greenhouse gases</td>
</tr>
<tr>
<td>HSP</td>
<td>Heat shock protein</td>
</tr>
<tr>
<td>HSPP</td>
<td>Heat shock protein promoter</td>
</tr>
<tr>
<td>LD</td>
<td>Lipid droplet</td>
</tr>
<tr>
<td>MIC</td>
<td>Minimum inhibitory concentration</td>
</tr>
<tr>
<td>MM+N</td>
<td>Minimal Medium + Nitrogen</td>
</tr>
<tr>
<td>MM-N</td>
<td>Minimal Medium - Nitrogen</td>
</tr>
<tr>
<td>OD&lt;sub&gt;600&lt;/sub&gt;</td>
<td>Optical Density (600 nm)</td>
</tr>
<tr>
<td>PA</td>
<td>Phosphatidic acid</td>
</tr>
<tr>
<td>PL</td>
<td>Phospholipid</td>
</tr>
<tr>
<td>POM</td>
<td>Pyruvate-oxaloacetate-malate (cycle)</td>
</tr>
<tr>
<td>PPP</td>
<td>Pentose phosphate pathway</td>
</tr>
<tr>
<td>SE</td>
<td>Sterol ester</td>
</tr>
<tr>
<td>SRS</td>
<td>Stimulated Raman scattering</td>
</tr>
<tr>
<td>TAG</td>
<td>Triacylglycerol</td>
</tr>
<tr>
<td>UNCTAD</td>
<td>United Nations Conference on Trade and Development</td>
</tr>
<tr>
<td>VLCFA</td>
<td>Very long chain fatty acid</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-type</td>
</tr>
<tr>
<td>YPD</td>
<td>Yeast Peptone Dextrose</td>
</tr>
<tr>
<td>YRBC</td>
<td>Yeast-recombination based cloning</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1| Energy densities of current and potential future mass-transport-energy sources.......................................................................................................................9

Figure 2| Overview of the main lipid biosynthesis pathways in *R. toruloides*... 19

Figure 3| Model of yeast lipid droplet formation and maintenance ..................25

Figure 4| Ascertaining cell volume using a brightfield image..........................39

Figure 5| Map of base vector pEGFP-Rt-YR-Hyg for transformation of *R. toruloides* ........................................................................................................................................44

Figure 6| Yeast Recombination-Based Cloning (YRBC) strategy for construction and verification of plasmids for organelle visualisation in *R. toruloides*........45

Figure 7| Construction by YRBC of *R. toruloides* GFP-tagged protein expression plasmids for organelle visualisation...............................................................46

Figure 8| Single slice confocal image of *R. toruloides* cells transformed with Ldp1-GFP ..................................................................................................................................48

Figure 9| Composite images of CARS/SRS and Ldp1-GFP tagged *R. toruloides* ..................................................................................................................................49

Figure 10| Comparison between raw confocal measurements of GFP-labelled Ldp1 organelles and the corresponding quantification of fluorescent intercellular compartments....................................................................................................................51

Figure 11| Rates of growth of GFP-Ldp1-transformed *R. toruloides* CBS14 with different glucose concentrations..................................................................................53

Figure 12| Median area of fluorescent regions within Ldp1-GFP-transformed *R. toruloides* ..................................................................................................................................54

Figure 13| Z-stack montage showing localisation of calnexin-GFP in an example *R. toruloides* cell..................................................................................................................55

Figure 14| Example image showing localisation of Atg8-GFP in *R. toruloides*. 57

Figure 15| Schematic of inducible promoters previously characterised in *R. toruloides* CBS14..................................................................................................................60

Figure 16| Strategy for generation of heat shock protein promoter (HSPP) fragment-containing reporter plasmid..................................................................................62
Figure 17 | PCR amplicons of heat shock protein promoters ............................... 64

Figure 18 | Example gel diagram of Yeast Colony-PCR verification of HSPPs. 65

Figure 19 | Representative gel of diagnostic plasmid digests of HSPP 19-26 66

Figure 20 | Induction of heat shock protein promoters controlled cell eGFP fluorescence .......................................................... 68

Figure 21 | Graphical series of measured fluorescence of heat shock protein promoters (HSPP) controlled eGFP after heat shock .............................. 73
1. Introduction

*Rhodotorula* (formerly *Rhodosporidium*) *toruloides* (Banno) is a heterothallic basidiomycete (subphylum Pucciniomycotina) yeast (Coelho et al., 2008; Wang, Q. M. et al., 2015) which has recently garnered interest as a biotechnologically relevant organism. *R. toruloides* strains have been isolated from many different ecological regions around the world: from the wood-pulp of Swedish conifers (Rennerfelt, 1937), to the air in the Koishikawa Botanical Garden, Japan (Okunuki, 1931), to the gut of a captive porpoise off of the Bahamas (Kutty and Philip, 2008). However, all these strains share common traits which are potentially useful in a biotechnological capacity: their ability to naturally synthesise carotenoids (useful as nutritional supplements), and their ability to accumulate neutral lipids at high density (with potential use in the petrochemical sector). Although other uses of the oleaginous yeasts – a term to describe organisms capable of greater than 20 % weight/weight oil accumulation (Evans and Ratledge, 1984a) – and their other pharmaceutically relevant products such as D-amino acids and L-phenylalanine ammonia-lyase have been alluded to in the literature (Liu, Y. et al., 2015; Park et al., 2018; Tully and Gilbert, 1985), the main focus of this thesis will be on developing tools to monitor lipid droplet biogenesis and its associated molecular machinery, intrinsically associated with neutral lipid accumulation, and identifying novel non-media-dependent promoters.

Neutral lipids, the primary constituent of lipid droplets in yeast, are rapidly manifesting themselves as an alternative to fossil fuels. Although the rate of growth of greenhouse gas (GHG) emissions appears to be finally slowing, as highly industrial countries move away from coal plants and the rate of renewable energy uptake increases (Jackson et al., 2016), climate change models still predict a likely anthropogenic climate change of between 1.6 °C to 3.2 °C above pre-industrial era temperatures from increasing atmospheric GHG (Rogelj et al., 2012). This is predicted to lead to corresponding catastrophic global effects as habitable land and food both deplete and extreme weather variability increases (Bathiany et al., 2018; Bebber et al., 2013; Jackson et al., 2016; Pacifici et al., 2015). As of 2018, the global production of oil refinery throughput is 8.2 million barrels day⁻¹, an increase of 2.1 % over the previous year (BP, 2004). Renewable energy sources have shown a marked production rise, with a global 17 %
increase above 2017 (excluding hydroelectric energy production) but this – while significant for static energy-drawing applications and for contributing to lowered coal and natural gas-derived carbon emissions – will be unable to effectively lower the oil-derived energy demand for long-distance travel (shipping and aeroplanes). Whereas an uptake of electric-powered vehicles can reduce individual anthropogenic GHG emissions, these vehicles currently rely on low-energy-density lithium-ion batteries or underdeveloped technology in hydrogen fuel-cells with an energy density of 2.64 MJ/L and 4.5 MJ/L at 690 bar respectively (Ulvestad, 2018), compared to 33.0 MJ/L for aviation gasoline and 40.8 MJ/L for high-sulphur marine fuel oil (Figure 1). In 2012 the vast majority of the year’s 961 million tonnes of CO₂ were released were from municipal marine vessels (1 806 650 vessels as of 2016). With >80 % of global trade by weight delivered through marine transportation (Walker, 2016) and an anticipated 50-250 % increase by 2050 (UNCTAD, 2016; Walker et al., 2018) it is apparent that, without a process by which these GHG are carbon neutral, the shipping and long-range transportation industry will have a marked impact on anthropogenic climate change.

![Energy densities of current and potential future mass-transport-energy sources.](https://www.energy.gov/eere/fuelcells/hydrogen-storage)
From the energy-density and predicted models mentioned above, it seems unlikely that the use of oil as the primary energy source of marine and aeronautical travel is likely to change in the next 50 years; prevention or reduction of further GHG will require an alternative source of the fuel. So-called ‘biofuels’ are carbon neutral; these sequester atmospheric CO$_2$ as they gain high-energy-density biomass before later releasing the GHG as they are combusted and, therefore, do not contribute to further GHG. ‘First generation’ biofuels have been developed successfully and these, while economical, are severely limited in their future application due to both the use of human-edible feedstocks, thereby bringing in ethical considerations as food sources become more in-demand, and bioethanol being limited to 5 % to 10 % anhydrous blending in conventional fossil fuels in unspecialised infrastructure in the EU and USA respectively (Charles et al., 2010; Naik et al., 2010; Tyner, 2010). In comparison, ‘second generation’ biofuels can use inexpensive, non-food biomass and, through generation of ‘drop-in’ biofuels, can be blended to 100 %, when in accordance with EN 14214:2012, and therefore reduce dependency on fossil fuels (‘Biodiesel Standards & Properties’).

*R. toruloides* is a potential second generation biofuel producer, natively capable of accumulation of lipids to over 65 % w/w (Zhang, Skerker, et al., 2016) to titres of 8 g L$^{-1}$ in shaking flasks; recent genetic overexpression of key lipid biosynthesis enzymes has raised this to 18.6 g L$^{-1}$ and to 89.4 g L$^{-1}$ lipid in fed-batch growth bioreactors (Zhang et al., 2016). Additionally, native *R. toruloides* is resistant to many lignocellulosic biomass hydrolysate-derived compounds released during the production of monomeric carbohydrates. The compound Acetic acid (up to 70 mM), a compound which strongly inhibits *Pichia stipitis* ethanol production (Delgenes et al., 1996), was found to be utilised as a lipid precursor and allowed accumulation of lipid content to 68 % w/w, though this is likely due to a shift to a carbon: nitrogen starved feedstock (Hu et al., 2009; Kitahara et al., 2014). The phenolic inhibitors 5-hydroxymethylfurfural (HMF), p-hydroxybenzaldehyde (PHB), syringaldehyde and vanillin did all significantly inhibit *R. toruloides* lipid biogenesis by between 5-15 %, although these were at concentrations far higher than found in actual biomass hydrolysates (Hu et al., 2009). Furfural, which leads to yeast cell-growth suppression and lipid biogenesis inhibition through sequestering of reductant NADH as the compound is reduced
to detoxified furfural alcohol (Nilsson et al., 2005), leads to a decrease in lipid content of *R. toruloides* by 26.5 % at concentrations 34 % and 10 % of that from spruce and corn stover hydrolysates respectively (Agbogbo and Wenger, 2007; Nilsson et al., 2005). However, mutant strains of the yeast isolated for tolerance to sugarcane bagasse hydrolysate, simultaneously developed greater tolerance to furfural with a minimum inhibitory concentration (MIC) of 5 g L\(^{-1}\); after a lag period of 36 hours (24 hours faster than the wild type (WT) *R. toruloides* mutant grew to a density comparable to non-furfural containing media containing 9.3 and 2.7 times that found in the spruce and corn stover hydrolysates at 2.5 g L\(^{-1}\) (Hu *et al.*, 2009; Kitahara *et al.*, 2014). As such, *R. toruloides* is capable of employing low-cost substrates. A 2016 pilot study by Dalmas Neto *et al.* using WT *R. toruloides* strain DEBB 5533 produced ASTM International standard biofuel when utilising sugarcane broth (40 g L\(^{-1}\) total sugars) supplemented with urea. Conversion of microbial oil to biodiesel was performed through base-catalysed transesterification using methanol and sodium hydroxide. Thliveros *et al.*, (2014) demonstrated production of fatty acid methyl esters using this transesterification method at a 97.7 % yield. The pilot study biodiesel had a total raw-material value of US$0.26 L\(^{-1}\) biofuel produced with an additional $0.50 L\(^{-1}\) biofuel cost associated with energy costs. With the price of crude oil at $0.34 L\(^{-1}\) in 2017 (Robson, 1981), this yeast-derived biofuel may be close to, or even capable of undercutting, the fossil-fuel prices if lower energy costs of lipid extraction can be realised. As reviewed by Xu and Liu, 2017, recent studies using low-value waste products such as crude glycerol (Polburee *et al.*, 2016) and domestic wastewater (Ling *et al.*, 2014) reported 69.5 % w/w and 63.95 % w/w lipid content respectively, although other traditional biomass hydrolysate substrates have also been utilised. *R. toruloides* CECT 13085 has been recently grown in increasing concentrations of non-detoxified wheat straw hydrolysates to selectively evolve a mutant strain on this low-value substrate capable of 39.5 g L\(^{-1}\) lipid yield when engineered to overexpress lipid pathway genes *DGAT1* and *SCD1*.

In addition to neutral lipid accumulation, which offers specific use as a biodiesel drop-in after transesterification to fatty acid methyl esters (FAMEs), *R. toruloides* has also been used to synthesise specific high-value lipids at high titres. After expression of galactose-inducible *Fusarium verticillioides* or *Mortierella alpina* \(\Delta 12\)-fatty acid desaturases (FADS) genes, overproduction of the essential fatty
acid linoleic acid increased fivefold to 1.3 g L⁻¹ (Wang, Y., et al., 2016). Similarly, integration of Δ12-fatty acid desaturase (FAD2) and ω3 desaturase (FAD3) from M. alpina and Aleurites fordii respectively also increased essential fatty acid α-linolenic acid (C18:3) production by a 4.5-fold increase in R. toruloides Δald1 (Lui et al., 2015). Introduction of the S. cerevisiae OLE1 gene, encoding Δ9-fatty acid desaturase, caused R. toruloides oleic acid (utilised in biolubricants, electrical insulation in transformers, and hydraulic fluid) content to increase fivefold, accounting for >70 % of total lipid (Park et al., 2018; Tsai et al., 2019). Exotic very long chain fatty acids (VLCFA) erucic acid (cis-docosa-13-enoic acid, C22:1Δ13) and nervonic acid (cis-tetracosa-15-enoic acid, C24:1 Δ15), important in production of plastics and as a treatment of human neurological pathologies such as Zellweger syndrome, adrenoleukodystrophy, schizophrenia, and multiple sclerosis, have also been produced to 10-15 g L⁻¹ (20-30 % of total fatty acid content). This has been achieved through ectopic heterologous integration of codon-optimised Brassica-plant species Crambe abyssinica and Cardamine graeca 3-ketoacyl-CoA synthases (KCS) (Amminger et al., 2012; Fillet et al., 2017; Tanaka et al., 2007).

Despite the recent acceleration of studies into R. toruloides, the molecular and genetic toolkit for R. toruloides remains limited; Agrobacterium tumefaciens-mediated transformation now allows relatively fast integration of DNA into the yeast, although codon optimisation is required for successful heterologous protein expression in this GC-rich organism, therefore necessitating custom synthesised oligonucleotides (Alexander Johns, pers. comm.). Targeted gene disruption has been demonstrated in Ku70 knockouts, required to facilitate weak homologous recombination, and CRISPR for the easily-identified CAR2 locus, encoding a key step in the carotenoid-biosynthesis pathway and therefore generating an albino colony when mutated (Koh et al., 2014; Liu, Y. et al., 2018; Otoupal et al., 2019).

1.1. Molecular process and organelle targeting

Having introduced the potential use of R. toruloides as a high-yield producer for drop-in biofuel production, the intracellular mechanisms and organelle properties of this yeast that facilitate such desirable properties will now be outlined.
1.1.1. Lipid biogenesis

Lipids are a group of organic molecules that are soluble in non-polar organic solvents and are requisite compounds for all living organisms. They are found in many diverse forms, each with differing properties, which fall within eight categories: fatty acids (FAs), glycerolipids, glycerophospholipids, sterol and sterol derivatives, sphingolipids, prenol lipids, glycolipids and polyketides (Fahy et al., 2011). Complex lipids are composed of an ester-linked carboxylic acid head group and poly(un)-saturated hydrocarbon tail(s) of various lengths. Yeast lipids can either be taken up from the cell's surroundings directly, produced by de novo synthesis, or recycled through the hydrolysis of existing lipids to synthesise the necessary new lipids; the following section reviews each of these processes and how it is optimised in R. toruloides for increased lipid accumulation.

External uptake of lipids from their surrounding media by yeasts is facilitated by either diffusion of lipids, or a temperature-dependent uptake step through the transporters Ypk1 and bifunctional Fat1 very long chain transporter (and acyl-CoA synthetase), and their subsequent activation by the energy-dependent acyl-CoA synthetase proteins Faa1/2/3/4 (Black and DiRusso, 2007; Jacquier and Schneiter, 2010). These activated FFA s (free fatty acids) can then be used as donors of acyl moieties during bi-acylation of the initial substrates glycerol or acyl-dihydroxyacetone phosphate with either the GPAT (glycerol-phosphate acyltransferase) proteins Sct1/Gpt2 or Ayr1 respectively to give lysophosphatidic acid (Lyso-PA). FFAs are then also used for acylation of Lyso-PA itself by AGPATs (acylglycerol phosphate acyl transferases) Slc1/4 or Loa1 to generate phosphatidic acid (PA). This complex lipid PA can then be used as the initial substrate in pathways for the biosynthesis of phospholipids: phosphatidylinositol (PI), phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), phosphatidylinositol (PS), and cardiolipin (CL).

As detailed in Figure 2, PA can, however, also be used as the base substrate in biosynthesis of neutral lipids. Cleaving inorganic phosphate from PA catalysed by the lipins Add1, Lpp1, App1, and (predominantly) endoplasmic reticulum (ER) membrane-bound Pah1 results in a diacylglycerol (DAG) molecule and this can then be acylated with the DGAT (diacylglycerol acyl transferases) Lro1 or Dga1 to triacylglycerols (TAGs) or to sterol esters (SEs) by the Are1/2 DGAT enzymes (Chae et al., 2012; Klug and Daum, 2014; Pascual and Carman, 2013; Soni et
In the ascomycete yeast *Saccharomyces cerevisiae*, deletion of *PAH1* significantly reduces the amount of DAGs available for acylation to TAGs thereby causing a proportional increase in cellular PL content; in stationary phase Δ*PAH1* mutants have a severely depleted lipid droplets (where TAG content is reduced by 80 %) and correspondingly high content of PLs (PC, +84 %; PE, +154 %; PI, +88 %; PS, +52 %; and PA; +450 %) and FFAs (+77 %) compared to WT (Fakas *et al.*, 2011). This lipid profile change leads to disruption of intracellular machinery through disruptive interactions and stresses, causing cell death through lipotoxicity (Listenberger *et al.*, 2003; Nguyen *et al.*, 2017; Wang, C. W., 2016). Curiously, *DGK1*-encoded diacylglycerol kinase, which reverses the dephosphorylation of PA to DAG back to DAG, restores the defect in lipid droplet formation back to WT but does not prevent the increased sensitivity of Δ*PAH1* mutants to lipotoxicity exhibited during growth in unsaturated FA-containing media (Fakas *et al.*, 2011).

However, biogenesis of complex lipids is not just limited to acylation from lipids taken up from media; precursor fatty acids are ubiquitously synthesised *de novo* in all yeasts for acylation of ADHP and glycerol-3-phosphate for feeding into the lipid storage or PL synthesis pathways (Klug and Daum, 2014). Oleaginous organisms (capable of lipid biogenesis greater than 20 % biomass) are particularly adept at lipid biogenesis through adaptations which overcome stoichiometric limitations in reducing power in the fatty acid synthase (FAS) enzyme complex for C<sub>16:0</sub>-C<sub>18:0</sub> synthesis from glucose; as each elongation cycle of the FAS complex requires two NADPH molecules, and one glycerol-derived acetyl-CoA, synthesis of C<sub>18:0</sub> is undertaken in 8 condensation reactions (Ratlledge, 2014). Therefore, for each mole of C<sub>18:0</sub> synthesised in the FAS complex 16 moles of NADPH are required. For each mole of glucose-derived acetyl-CoA in glycolysis, however, only 9 moles of NADPH are generated; oxidising cofactors generated in the mitochondria are assumed to be used almost exclusively in ATP synthesis (Ratlledge, 2014). The pentose-phosphate pathway (PPP) is a second cytoplasmic pathway capable of producing an additional 6 moles of NADPH directly, but also generates an additional mole of NADPH through feeding of PPP-derived glyceraldehyde-3-phosphate into the pyruvate-oxaloacetate-malate (POM) cycle; this would satisfy the stoichiometric requirements for *de novo* FA biosynthesis. For simplicity, to assume only C<sub>18:0</sub>
acyl groups are present in synthesised TAGs and two ATP molecules generated for each terminal oxygen in the mitochondrial electron transport chain, a theoretical yield of 31.6 g of TAG can be synthesised from 100 g of glucose (Klug and Daum, 2014; Ratledge, 2014). In the oleaginous yeast *Yarrowia lipolytica* without overexpression of cytosolic malic enzyme, overexpression of the NADP reducer is reported by Macool *et al.*, 2013 to increase lipid content by 5-10%.

For simplicity, *de novo* TAG detailed hereafter will use only glucose as an original carbon source. Cytoplasmic glycolysis-synthesised 3-carbon dihydroxyacetone phosphate (DHAP) can either be: acylated, as detailed above, with the glycerol-phosphate acyl transferases Sct1 (GPAT2) or Gpt1 (GPAT2) to acyl-dihydroxyacetone phosphate; dehydrogenated with the NAD+-dependent glycerol-3-phosphate dehydrogenase Gpd1 to glycerol-3-phosphate; or alternatively continue through glycolysis to generate 3-carbon pyruvate (Klug and Daum, 2014; Zheng and Zou, 2001). Passing through the mitochondrial membranes using the pyruvate decarboxylase complex, 2-carbon acetyl-CoA is generated and can be used to re-form 6-carbon citrate from 4-carbon oxaloacetate in the Krebs cycle; this citrate is crucial in facilitating oleaginous yeast lipid assimilation. Under lipid-accumulating nitrogen starvation conditions *IDH1/4*-encoded isocitrate dehydrogenase enzyme, responsible for the catalysis of isocitrate to oxoglutarate in the third step of the Krebs cycle, is negatively regulated leading to an increase in isocitrate and, due to the reversible nature of the aconitase enzyme, an increase in citrate in the mitochondria (Ratledge, 2014; Shi, S. and Zhao, 2017; Tang *et al.*, 2009; Yang *et al.*, 2012). This citrate can be transported to the cytoplasm through, firstly, the mitochondrial citrate transport protein across the inner membrane, before passively diffusing through a selective ion channel. Once in the cytoplasm, citrate is hydrolysed to acetyl-CoA and oxaloacetate through the action of AclI ATP-citrate lyase with the acetyl-coA feeding into the FAS complex and oxaloacetate being oxidised to malate for re-transporting back into the mitochondria and use in the Krebs cycle. This oxaloacetate, in some oleaginous yeast including *R. toruloides*, is able to instead be used to oxidise malate through the POM cycle-facilitated ‘malic enzyme’ to coupled-reaction reduce NADP to NADPH (Evans and Ratledge, 1984a; Macool *et al.*, 2013; Ratledge, 2014; Zhang, Ito, *et al.*, 2016).
De novo synthesis of FFAs in euakarytes occurs in both the cytoplasm/ER and the mitochondria. Cytoplasmic/ER based synthesis is the principal site of TAG synthesis for lipid droplet formation and is the focus of this section. In an ATP-dependent reaction catalysed by a biotinylated acetyl-CoA carboxylase Acc1, pyruvate-derived acetyl-CoA is carboxylated to a malonyl-CoA three-carbon building block (Choudhary et al., 2018; Klug and Daum, 2014; Mishina et al., 1980; Wang, H. et al., 2017). FA synthesis then occurs in the FA synthetase (FAS) complex; with each ‘cycle’ of the FAS functionally conserved proteins elongate the nascent acyl group by two carbons.

In the ascomycete S. cerevisiae, a giant multifunctional 2.6 MDa αβ6 heterododecamer of the proteins 230 kDa Fas1 and 210 kDa Fas2 contains a central hexameric α wheel capped by two trimeric β domes for transfer of two carbons from malonyl-CoA to a saturated hydrocarbon chain (Leibundgut et al., 2007). The S. cerevisiae Fas1 β-subunit contains the highly mobile ‘acyl carrier protein’ (ACP) acetyl transferase, enoyl reductase, dehydrase 1 & 2 and the larger part of a malonyl-palmitoyl transferase domain. The Fas1 α-subunit provides the remainder of the malonyl-palmitoyl transferase, acyl-carrier proteins, 3-ketoreductase, 3-ketosynthase, and phosphopanthetheine transferase. In standard growth conditions, the FFAs produced by the cytoplasmic FAS are saturated C\textsubscript{16} palmitic acid or C\textsubscript{18} stearic acid molecules, synthesised in a 2:3 ratio (Fischer et al., 2015; Leibundgut et al., 2007; Schweizer, E. and Hofmann, 2004; Schweizer, M. et al., 1986).

Interestingly, the R. toruloides 2.7 MDa FAS type 1 protein complex contains an evolutionarily recent ACP duplication event (Fischer et al., 2015). As both ACP domains show 77 % sequence identity in their mobile linker regions, with an especially high homology in the acyl-docking sequence, it is thought that these seemingly identically-functional protein subunits work to only increase relative crowding of the enlarged R. toruloides FAS (18 Å wider, though 20 Å shorter than the barrel found in in S. cerevisiae) with substrates, and this facilitates a faster reaction rate due to a localised increased substrate concentration. Gene splitting of the FAS 1 complex also appears to have occurred in the anti-parallel β-sheet between the enoyl reductase and dehydrase domains of the complex though this is not hypothesised to have any functional effect (Fischer et al., 2015).
Fatty acids are activated to functional fatty acid-CoA (FA-CoA) by the synthases Faa1/4 or a Fat1; these activated molecules can then be used as acyl-donors directly in the synthesis of lyso-PA and TAGs by the GPAT or DGAT enzymes as previously mentioned, or further elongated or desaturated in the ER (Oh et al., 1997; Schweizer, E. and Hofmann, 2004; Tehlivets et al., 2007).

Elongation and desaturation of FAs occurs in the ER using the elongase enzymes Elo1, Elo2, and/or Elo3 – to add additional malonyl-CoA carbons – and desaturation enzyme Ole1, to generate the required lipids (Oh et al., 1997; Tehlivets et al., 2007). Elo1, Elo2 and Elo3 facilitate sphingolipid biogenesis, in conjunction with accessory proteins beta-keto reductase Ybr159w and 3-hydroxyacyl Co-A dehydrase, as part of the FA elongation cycle (Browne et al., 2013). Whereas Elo1 elongates C\textsubscript{14}-C\textsubscript{16} to C\textsubscript{16}-C\textsubscript{18} medium length FAs, its paralogues Elo2 and Elo3, derived from the whole-genome duplication of \textit{S. cerevisiae}, elongate FAs shorter than C\textsubscript{22} up to C\textsubscript{24}, and elongate C\textsubscript{24} FAs to C\textsubscript{26} respectively (Oh et al., 1997). As such, Elo2 and Elo3 are necessary for very long chain fatty acid (VLCFA) synthesis and their mutants show marked differences in sphingolipid composition in affected cells, thereby showing membrane generation defects. Released VLCFA is used as the substrate in AGPAT synthesis to facilitate position-2 acylation of lyso-PA (Benghezal et al., 2007; Klug and Daum, 2014).

Because an excess of free cytoplasmic or ER-localised FAs from any of the three FA biosynthesis pathways (such as through inoculation in oleic-acid containing media) can disrupt cellular machinery and cause lipotoxicity, they are rapidly either: used in the synthesis of complex lipids; stored for later use as SE or TAGs in the lipid droplet; or translocated into peroxisomes for catabolic β-oxidation by the ABC transporter heterodimers Pxa1-Pxa2 (esterified long chain FAs such as oleic acid C\textsubscript{18}:1) or Ant1-Pex11 (non-activated medium chain FAs) before activation by the aforementioned acyl-CoA synthetase Faa2 (Baker et al., 2015; Klug and Daum, 2014; Palmieri et al., 2001).
Figure 2| Overview of the main lipid biosynthesis pathways in *R. toruloides*. A summary of the biosynthetic pathways in yeast, adapted from a combination of figure 2 and figure 3 in Klug and Daum, 2014 and figure 3 in Ratledge, 2014. Nitrogen starvation response is highlighted in purple; abbreviated pathways are shown by dotted arrows and named in orange; POM cycle is indicated by green arrows; enzymes are shown in red and their gene names and functions in *S. cerevisiae* are highlighted in table 1; genes that have been experimentally overexpressed, introduced from foreign organisms or deleted in *R. toruloides* are in a light blue box; co-factors and by-products are labelled in orange text (Gro, glycerol; CoA, coenzyme A); lipid-biosynthesis products and intermediates are shown in black text. Where an enzyme shows activity in more than one organelle then the site where biosynthesis activity is highest is shown. The yellow-bordered area represents lipid droplets; the orange-bordered area represents the mitochondria; the green-bordered area represents the endoplasmic reticulum and the red-bordered white area represents the cell membrane and cytoplasm respectively.
Table 1 | Key genes identified in oleaginous eukaryotic lipid biosynthesis.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Gene name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acc1</td>
<td>Acetyl-CoA Carboxylase</td>
</tr>
<tr>
<td>Acl1</td>
<td>ATP citrate lyase</td>
</tr>
<tr>
<td>Aco1</td>
<td>Aconitase</td>
</tr>
<tr>
<td>Ald1</td>
<td>Cytosolic aldehyde dehydrogenase</td>
</tr>
<tr>
<td>App1</td>
<td>Phosphatidate phosphatase</td>
</tr>
<tr>
<td>Are1/2</td>
<td>Acyl-CoA:sterol acyltransferase</td>
</tr>
<tr>
<td>Ayr1</td>
<td>1-Acyl-dihydroxyacetone phosphate reductase</td>
</tr>
<tr>
<td>Cit1</td>
<td>Citrate synthase</td>
</tr>
<tr>
<td>Dga1</td>
<td>Diacylglycerol acyltransferase</td>
</tr>
<tr>
<td>Dgk1</td>
<td>Diacylglycerol kinase</td>
</tr>
<tr>
<td>Dpp1</td>
<td>Diacylglycerol pyrophosphate phosphatase</td>
</tr>
<tr>
<td>Elo1</td>
<td>Elongase I</td>
</tr>
<tr>
<td>Faa1</td>
<td>Long chain fatty acyl-CoA synthetase</td>
</tr>
<tr>
<td>Faa4</td>
<td>Long chain fatty acyl-CoA synthetase</td>
</tr>
<tr>
<td>Fad2</td>
<td>Δ12-fatty acid desaturase</td>
</tr>
<tr>
<td>Fad3</td>
<td>ω3 desaturase</td>
</tr>
<tr>
<td>Far</td>
<td>Fatty acid reductase</td>
</tr>
<tr>
<td>Fat1</td>
<td>Very long chain fatty acyl-CoA synthetase and fatty acid transporter</td>
</tr>
<tr>
<td>Fit2</td>
<td>Facilitator of Iron Transport</td>
</tr>
<tr>
<td>Gpt1</td>
<td>Glycerol-3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>Gpt2</td>
<td>Glycerol-3-phosphate acyl transfer</td>
</tr>
<tr>
<td>Idh1/4</td>
<td>Isocitrate dehydrogenase</td>
</tr>
<tr>
<td>KCS</td>
<td>3-ketoacyl-CoA synthases</td>
</tr>
<tr>
<td>Loa1</td>
<td>Lysophosphatidic acid acyltransferase</td>
</tr>
<tr>
<td>Lpp1</td>
<td>Lipid phosphate phosphatase</td>
</tr>
<tr>
<td>Lro1</td>
<td>Lecithin cholesterol acyl transfer Related Open reading frame</td>
</tr>
<tr>
<td>M. E.</td>
<td>Malic Enzyme</td>
</tr>
<tr>
<td>Mdh1</td>
<td>Malate dehydrogenase</td>
</tr>
<tr>
<td>P. D. C.</td>
<td>Pyruvate dehydrogenase complex</td>
</tr>
<tr>
<td>Pah1</td>
<td>Mg2+-dependent phosphatidate phosphatase</td>
</tr>
<tr>
<td>Pdc1</td>
<td>Pyruvate decarboxylase isozyme</td>
</tr>
<tr>
<td>Pyc1/2</td>
<td>Pyruvate decarboxylase</td>
</tr>
<tr>
<td>Scf1</td>
<td>Suppressor of Choline Transport mutants</td>
</tr>
<tr>
<td>Slc1/4</td>
<td>1-acyl-sn-glycerol-3-phosphate acyltransferase</td>
</tr>
<tr>
<td>Tgl3</td>
<td>Triacylglycerol Lipase</td>
</tr>
<tr>
<td>Tgl4</td>
<td>Triacylglycerol Lipase</td>
</tr>
<tr>
<td>Tgl5</td>
<td>Triacylglycerol Lipase</td>
</tr>
<tr>
<td>Tsc13</td>
<td>Enoyl reductase</td>
</tr>
<tr>
<td>Yju3</td>
<td>Monoglyceride lipase (MGL)</td>
</tr>
</tbody>
</table>
1.1.2. Lipid droplets

Lipid droplets (LDs) are intracellular organelles found ubiquitously in all eukaryotic, most bacterial, and some archaeal cells although their constituent lipid molecule makeup and indeed the names of these organelles are not well conserved; oil bodies, oleosomes, granules, adiposomes, lipid droplets and lipid bodies are all terms used for different taxa. Most bacteria and all archaea synthesise polyhydroxyalkanoates and their derivatives as their main lipid droplet constituents, in contrast eukaryotes (and TAG-biosynthesis capable prokaryotes) synthesise lipid droplets which contain predominantly neutral-lipid TAGs and SEs (Brandl et al., 2005; Murphy, 2012).

First described in the late 19th century, LDs were considered relatively static carbon-storage organelles with a low rate-of-turnover, for much of the 20th century before a rapid acceleration in their study commenced in the 1990s, spurred by their apparent role in mammalian pathologies including atherosclerosis, obesity, dyslipidaemia and cancers (Xu, S. et al., 2018). Now known to be multifaceted dynamic organelles, LDs are implicated not only in lipid storage but also in cell signalling, protein maturation, protein storage, resistance to ER and oxidative stresses by environmental insults, membrane formation, and the storage of vitamins precursors. Lipid droplet dynamics and investigations into these organelles and their intracellular interactions are now a component in most avenues of eukaryotic cellular investigations. However, it is lipid droplet biosynthesis, maintenance, and degradation processes and regulation that are most crucial in the context of R. toruloides as a putative oleaginous yeast for biofuel production.

As outlined in section 1.1.1, the terminal location of lipid synthesis is the ER where diacylglycerol phosphate acyl transferases (DGATs) Lro1 and Dga1 are embedded and enriched within the inner and outer leaflets respectively, forming punctate regions of TAGs within a lens. It is then thought that the barrel-shaped seipin homo-oligopeptide scans the ER bi-layer membrane until a protrusion is encountered where neutral lipids have spontaneously accumulated because of the lower free energy cost between TAGs than between TAGs and PL (Liu, X. N. et al., 2016; Sui et al., 2018; Thiam and Forêt, 2016). Through unknown means, this punctate region then anchors and recruits transmembrane lipid phosphatase Fit2 in the ER membrane, which specifically binds to TAGs, and GPATs, which
catalyse the rate limiting step in TAG biosynthesis as shown in Figure 3. It is thought that these proteins both structurally support the inner leaflet of the nascent LD (<200 nm diameter), thereby preventing invaginations of the LD into the ER lumen rather than the cytosol, and also constricting to provide a neck to the LD for regulation and maintenance of the ER/LD interface. In Drosophila melanogaster S2 cells, seipin deletion mutants are unable to develop morphologically-normal pre-LDs (300–500 nm diameter) and, instead, remain as relatively mobile puncta in the ER (Liu, X. N. et al., 2016). Additionally, enrichment of negative molecular-curvature glycerolipids is found at the ER/LD interface site and mutations of FIT2 and deactivation of PAH1 by deletions of PAH1 phosphatases NEM1 and SPO7, which dephosphorylate phosphatidic acid to the negative-curvature diacylglycerol lipid, prevent LD emergence; this finding is consistent with computational models of ER/LD tubule development where increased convex lipids lead to spontaneous LD emergence (Choudhary et al., 2018; Garay et al., 2014; Miranda et al., 2014).

As summarised by Schuldiner and Bohnert, (2017a), the mature LD, having a hydrophobic neutral core encapsulated by a PL monolayer, is prevented from ‘normal’ intracellular shuttling between aqueously-luminal organelles as found in fusion and fission of bilayer organelles. In contrast to mammalian cells, where LDs are able to either entirely disassociate from the ER or maintain contact through ER/LD bridges, mediated through a Rab18-NAG-RINT1-ZW10 (NRZ)-SNARE complex, yeast species maintain contact between the ER and LD at all times (Schuldiner and Bohnert, 2017; Wilfling et al., 2013; Xu, D. et al., 2018). These LD/ER contact sites help prevent ER stress and sequester cytotoxic lipid and proteins in both yeast and mammals that develop during metabolic and environmental stresses. As outlined in section 1.1.1, lipotoxic FAs are sequestered and detoxified to neutral TAGs by DGAT1 in the ER; in cells where LD biogenesis is inhibited an increased sensitivity to toxic lipids can be observed (Klecker et al., 2017; Listenberger et al., 2003; Nguyen et al., 2017). In mammals, LDs have also been implicated in both clearance of non-functional proteins where ubiquitinated apolipoprotein B accumulates and aggregates before degradation by autophagy (Ohsaki, 2006), and in human pathologies where α-synuclein aggregates with LDs in Parkinson’s disease (Colebc et al., 2002). Proteotoxic-temperature derived misfolded proteins, similarly, aggregate in yeast, termed
‘inclusion bodies’ (IB). In *S. cerevisiae* these IB were found by Moldavski *et al.*, (2015) to contain 13 resident proteins, of which six are crucial for IB clearance after stress. Having found IB and LD to co-localise during stress (Szymanski *et al.*, 2007), Moldavski *et al.*, (2015) further demonstrated an interaction between the two organelles, facilitated by the IB protein ‘increased minichromosome loss’ (Iml1). Iml1 demonstrates a binding interaction with LD proteins Pet10 (Plin1) and Pdr16, with a preferential binding at 37 °C. The group then suggested a pathway by which the IB was cleared that involved LDs. Δdga1 Δlro1 double mutants, where LDs are unable to synthesise TAGs, were able to clear the IB but Δare1 Δare2 double mutants, where LDs cannot synthesise sterol esters, cannot. LD-constituent neutral-lipid sterol derivatives may therefore act as a ‘chemical chaperone’ and solubilise hydrophobic interactions between integrated aggregates of IB proteins to facilitate refolding and disaggregation of the aberrant proteins.

Peripilin class proteins, which are able to sterically prevent spontaneous TAG lipolysis by preventing access by cytoplasmic lipases and fusion of established LDs, also associate to the LD outer leaflet from the cytoplasm; in yeast, the *S. cerevisiae* protein peripilin Plin1 quickly co-accumulates nascent-LDs. In common with other peripilin homologues in mammals and plants, the peripilin is thought to help anchor the PL-monolayer to the neutral-lipid core (Gao *et al.*, 2017). Peripilins may also help facilitate vectorial ER-membrane deformation towards the cytoplasm (Henne *et al.*, 2018a) because of the capacity of Plin1 mammalian-homologue Plin3 to spontaneously reorganise liposome membranes, similar to apolipoproteins (Bulankina *et al.*, 2009).

As carbon source concentrations reduce in the yeast medium, *S. cerevisiae* LDs associate with the nucleus, where Pah1 proteins also begin to accumulate. As such, during a diauxic shift, the TAG precursor diacylglycerol is increasingly synthesised, coinciding with the yeast starvation response of increased lipid droplet size (Schuldiner and Bohnert, 2017). However, as the yeast moves towards its starvation response, LDs disassociate from the nucleus and localise to sterol-rich lipid rafts on the vacuole; if starvation continues the LDs become engulfed by the vacuole and macrolipophagy occurs as TAGs are catabolised to FFAs (van Zutphen *et al.*, 2014). Additionally, FFAs can be liberated from TAG storage molecules through the activities of TAG lipases Tpl3, Tpl4, and Tpl5;
which are localised to LD outer membranes (Athenstaedt and Daum, 2003). Curiously, Tpl3 function is confounded by both a lipase and lysophosphatidic acid acylation function, though this is likely to assist with membrane stability (Rajakumari and Daum, 2010).

In S. cerevisiae, Atg8 protein can also be recruited to degrade lipid for FFA recycling. This occurs through microautophagy, although the direct mechanism for this process is yet to be elucidated. Briefly, Atg8 recruits additional Atg proteins to form a de novo preautophagosomal structure (PAS) which associates with peripilin-family proteins, the PLINs. This structure then appears to sequester and engulf small LD volumes in a piecemeal fashion (Schulze et al., 2017; van Zutphen et al., 2014).
Figure 3 | Model of yeast lipid droplet formation and maintenance. Initially, seipin oligopeptides scan the ER leaflets for a membrane disruption and deformation as neutral lipid accumulations form an intra-membrane lens. This then stabilises and anchors the LD in place, and recruits proteins associated with TAG and negative-curvature phospholipid-membrane biosynthesis, as indicated by grey triangles; positive-curvature phospholipids are represented in black triangles. Additionally, lipid droplets are protected from spontaneous lipolysis by sterically inhibiting peripilins, represented by blue circle with membrane-embedded tails. Mammalian macrolipophagy, whereby Atg8/LC3 protein in mammals nucleates and develops a double-membrane autophagosome, envelopes whole LDs and targets them for lipophagy-lipid recycling as lysosomes fuse to the nascent autophagolysosome. Figure is a compilation of many figures including (Gao et al., 2017; Liu, X. N. et al., 2016; Sui et al., 2018; Szتلryd and Brasaemle, 2017; Szymanski et al., 2007).
1.2. Heat shock promoters

A significant amount of work has recently been undertaken to improve and increase the genetic tools available for *R. toruloides* strains.

To date, the unreproducible method of stably introducing desired DNA constructs into the yeast through polyethylene glycol-facilitated protoplast transformation (Tully and Gilbert, 1985) has been superseded by *Agrobacterium tumefaciens*-mediated transformation (ATMT) (Johns *et al.*, 2016; Lin *et al.*, 2014; Liu, Y. *et al.*, 2013). Recent studies have also reported the successful use of electroporation and lithium acetate-mediated transformation (it may be noted, however, that I have been unable to replicate the findings of Lui *et al.*, 2017, data not shown) (Liu, H. *et al.*, 2017; Otoupal *et al.*, 2019; Takahashi *et al.*, 2014; Tsai *et al.*, 2017). Endogenous gene knockouts have also recently been reported in *R. toruloides* strains: Koh *et al.* (2014) achieved disruption of the *CAR2* gene in a Δ*ku70* *R. toruloides* ATCC 10657 strain through the use of homologous gene flanking sequences and ATMT; *R. toruloides* DMKU3-TK16 *URA3* was successfully disrupted using an optimised lithium acetate-mediated transformation by Tsai *et al.*, (2017); and single and multiplex disruption of both *URA3* and *CAR2* genes were achieved in *R. toruloides* IFO 0880 (ATCC 10657) by CRISPR, with an editing efficiency of 3.4 % ±2.7 % to 46.2 % ±22.2 % for each of the genes respectively (Otoupal *et al.*, 2019).

The genetic toolkit in *R. toruloides* also consists of eight identified native constitutive promoters (Liu, Y. *et al.*, 2013, 2016; Wang, Y., Lin, *et al.*, 2016), one heterologous promoter from the basidiomycete *Ustilago maydis* (Liu, Y. *et al.*, 2018), with a further seven native inducible promoters, inducible or repressible under nutrient-limited conditions, having been characterised (Johns *et al.*, 2016; Liu, Y. *et al.*, 2015). However, these inducible promoters are limited in their applications in a commercial setting for synthesis of specialty chemicals where either the feed is not fully controllable (as in production of lipid and lipid-derived compounds from low-cost substrates as summarised by Xu and Liu, 2017) or where limiting the growth medium to induce promoter-driven requisite genes may cause disadvantageous metabolic remodelling (Broach, 2012; Cai *et al.*, 2011; Dechant and Peter, 2008; Klosinska *et al.*, 2011). Consequently, there is a need
for identification and development of strong, tightly controlled inducible promoters not dependent on media chemical-makeup to fulfil this biotechnological role.

One such candidate for gene induction is the environmental stress response (ESR) in yeast, where the yeast genome expression profile is dramatically altered to resist and accommodate potentially damaging physiological conditions by modulating metabolic and biosynthetic pathways. Gasch et al. (2000) found approximately 900 genes whose transcription levels were rapidly and significantly decreased (repressed) or increased (induced) in response to a variety of ESRs; a flurry of microarray studies in the early 2000s of global expression profiles under various environmental stresses included: temperature shocks (Causton et al., 2001; Gasch, A. P. et al., 2000); oxidative stresses (Gasch, A. P. et al., 2000); ethanol shocks (Alexandre et al., 2001); starvation of amino acids, nitrogen, in stationary phase (Gasch, A. P. et al., 2000); and diauxic shifts (Gasch, A. P. et al., 2000), amongst others. While some genes were almost universally either up- or down-regulated after environmental stress, it was apparent that an environmental assault does not alter all genes’ transcription equally or consistently after differing assaults.

The ESR is orchestrated through a plethora of specific and general transcription factors which can antagonistically or concomitantly affect global mRNA transcription levels of cis-regulated proteins. The mechanisms of these transcription factors are highly conserved throughout the eukaryotes, and indeed the DNA-binding domain of key eukaryotic heat-shock factor Hsf1 with bacterial heat-shock sigma factor σ32 which facilitates RNA polymerase binding and transcription in bacteria (Trott and Morano, 2007). After an environmental assault, transient genetic remodelling occurs with an average relative-mRNA level shift reaching 100 % after 2-5 minutes, and reaching peak expression after 15 minutes (Causton et al., 2001). In the 900 genes in S. cerevisiae affected by the ESR, the physiological changes can be attributed primarily to an alteration in gene expression of proteins controlling carbohydrate metabolism, oxidative stresses, protein folding and targeted degradation, autophagy, cytoskeletal rearrangement and intracellular signalling to mitigate the cytotoxic effects of the environment the cell suddenly encounters.
After an assault by many different conditions, the carbohydrate composition of yeast cells is rapidly altered. Originally thought to exist primarily in spores, trehalose, the disaccharide (α-D-glucopyranosyl-1,1-α-D-glucopyranoside), changes composition from 6.53 to 13.66 mg g⁻¹ dry weight, and there is an alteration of 40.0 to 52.0 mg g⁻¹ in total fatty acid, after a heat shock from 21 °C to 37 °C for 60 minutes in S. cerevisiae 3001 (Ballio et al., 1964; Odumeru et al., 1993; Shi, L. et al., 2010). This trehalose carbohydrate has been implicated in cellular protection in many ways. Principally, trehalose has been shown to quench oxygen radicals which become overrepresented in heat-shocked cells (Benaroudj et al., 2001; Weng and Elliott, 2015), a function critical to cell viability as an overabundance of reactive oxygen species (ROS) leads to protein crosslinking and fragmentation as protein backbones are oxidised (Berlett and Stadtman, 1997), phospholipid membrane and lipid-signalling damage through peroxidation (Ayala et al., 2014), and macromolecular sugars alteration (including DNA-backbone damage) (Rowe et al., 2008). Mutants of the trehalose-subunit synthase genes TPS1 and TPS2 lose their thermotolerance (De Virgilio et al., 1994), a phenotype which can be rescued by supplemental extracellular trehalose introduction. In addition to these antioxidant properties, trehalose also prevents cytotoxic protein accumulation from heat shock-induced peptide misfolding and maintains protein activity when cells where trehalose is depleted are non-functioning, although the mechanism of this thermo-stabilisation appears to be multifaceted and is not yet fully elucidated (Benaroudj et al., 2001; De Virgilio et al., 1994; Singer and Lindquist, 1998; Sola-Penna et al., 1997).

Curiously, trehalose synthase genes TPS1 and TPS2, and the neutral trehalose degradation gene NTH1, are all co-induced by the ESR. However, during heat shock at 40 °C, S. cerevisiae S288c trehalase is phosphorylated and deactivated whereas the trehalose synthases become three times more active at this temperature compared to 25 °C, implying that protein kinetics also affects the physiological ESR (Neves and François, 1992). In R. toruloides strain CBS 14, Evans and Ratledge (1984b) found 6.2-6.5 % of dry biomass is trehalose when cultured in nitrogen-limited medium containing 30 g L⁻¹ glucose.

In the same way that an increase in temperature can cause an increase in trehalose synthase activity, a heat shock assault will alter the protein kinetics and conformational and interactive properties of all unprotected proteins and
enzymes within a cell. Non-functional protein conformers can be produced where environmental stress causes intermediate misfolded-peptides which are unable to escape their free-energy minima (Jahn and Radford, 2005). In such conformations toxic misfolded proteins often expose hydrophobic surfaces normally hidden in a protein hydrophobic core and co-aggregate (Vabulas et al., 2010), before either refolded or becoming ubquintated and terminally-aggregated; aberrant proteins are then actively sequestered in inclusion bodies and targeted for degradation (Ohsaki, 2006; Tyedmers et al., 2010). Lipid droplets are also implicated in detoxifying inclusion bodies through a sterol derivative, as outlined in section 1.1.2.

In addition to the carbohydrate remodelling touched on above, where trehalose is capable of detoxifying or sequestering some of the increased ROS from intra- and extracellular sources found in an environmental assault, genes encoding proteins specifically for this role are upregulated in S. cerevisiae. With this purpose the protein-coding genes of thioredoxin (TRX2), glutaredoxin (TTR1) and glutathionine (ECM38) isozymes (oxidoreductases maintain the internal oxidative and reductive-molecule homeostasis in yeast through reduction of NADPH) (Grant, 2001), antioxidants (PRX1), and cytosolic superoxide dismutase (SOD1) are upregulated (Ayer et al., 2014; Causton et al., 2001; Gasch, A. P. et al., 2000; Gasch, Audrey P., 2007); these help prevent unmanageable oxidation of yeast macromolecules and subsequent cell death. Furthermore, lipid droplets have been associated with protection against cytotoxic ROS by sequestering the lipophilic radicals (Herms et al., 2013). In vivo, Drosophila melanogaster glial cells induce LD accumulation during ROS insult, and correspondingly reduce LD accumulation during low ROS, possibly implying an inducible protective effect of LD against ROS (Liu, L. et al., 2015).

These ESRs to heat shock are modulated by the conserved Msn2/4 general stress response and in invertebrate animals and yeast the transcriptional activator Hsf1; in mammals there are four HSF family members (Åkerfelt et al., 2010). Ascomycete S. cerevisiae Msn2 and Msn4 transcription factors contain a zinc-finger domain which bind to the highly conserved stress response element (STRE) CCCCCT (Schmitt and McEntee, 1996; Underhill, 2012) which, upon an environmental assault, rapidly translocates from the cytoplasm to the nucleus (Görner et al., 1998). Msn2 protein relocations to the nucleus are mediated by
hyperphosphorylation of a 250-residue sequence and subsequent disassociation from the 14-3-3 family cytosolic adapter protein Bmh2 (Beck and Hall, 1999; Boy-Marcotte et al., 2006) where several hundred are down-regulated and 100-200 genes are up-regulated respectively through the binding of the cis-acting STRE sequence (Causton et al., 2001; Gasch, A. P. et al., 2000). Msn2 and Msn4 activity is then attenuated by either nuclear proteasomal degradation or returning to the cytoplasm, primarily through the Msn5 nuclear exportin (Bose et al., 2005; Chi et al., 2001; Durchschlag et al., 2004).

Yeast heat shock factor 1 (Hsf1) differs from the metazoan homologue in that it constitutively remains as a DNA-bound trimer (metazoans undergo a monomer-to-trimer transition when activated), having bound to stress response elements (STE) prior to an environmental assault (Gross et al., 1990). Upon heat shock, however, this basally constitutive trimer transcriptional activity increases ~20 fold as adjacent HSF trimers appear to complex together to “unmask” activator domains and complex together as they are post-translationally phosphorylated (Lee et al., 2000). Additionally, an accumulation of denatured proteins, derived from heat shock assault as described above, can indirectly activate yeast Hsf1 (Moraitis and Curran, 2004; Parsell and Lindquist, 2003; Trotter et al., 2002). There appears to be an intrinsic interplay between both the general stress response and the heat-specific stress responses, where many HSP gene promoters contain both STREs and STEs (Amorós and Estruch, 2001).

1.3. Project aims

*R. toruloides* is an oleaginous yeast with potential use as a biotechnological chassis for production of biofuel and lipid droplet-derived products such as bio-lubricants, bio-plastics and pharmaceutically relevant lipids as outlined in sections 1.1 and 1.2. However, before the yeast can be used to synthesise these desired products, further genetic tools and a greater understanding of the underlying mechanisms of lipid accumulation must be realised. To this end, this project aims to develop and begin characterisation of genetic tools for controlled induction in undefined media, through heat-shock inducible promoters, and develop tools to facilitate further investigations into lipid droplet assembly and accumulation through live-cell imaging of relevant organelles.
2. Materials and Methods

2.1. Strains and growth conditions

2.1.1. *Rhodotorula toruloides*

Wild-type haploid *R. toruloides* CBS 14 (MAT-A1; ATCC 10788; ATCC 15385; CCRC 20306; DBVPG 6740; IAM 13469; IFO 0559; IGC 4416; JCM 3792; MUCL 30249; NCYC 921; NRRL Y-1091; VKM Y-334; PYCC 4416) (BANNO, 2008; Rennerfelt, 1937) was acquired from the Westerdijk Fungal Biodiversity Institute (formerly the Centraalbureau vor Schimmelcultures).

*R. toruloides* was grown: initially at 25 °C in 6.9 g L⁻¹ Yeast Nitrogen Base without Amino Acids (YNB) (Formedium, Hunstanton) with 790 mg L⁻¹ Complete Supplement Mix (CSM) (Formedium, Hunstanton) and 20 g L⁻¹ glucose before temporary heat shock at 37 °C; constitutively in YNB and CSM with a glucose concentration of 0 g L⁻¹ to 20 g L⁻¹ during glucose starvation studies; or as described in Zhu et al., (2012) on Minimal Media with Nitrogen (MM+N) (25 g L⁻¹ glucose, 0.5 g L⁻¹ MgSO₄·7H₂O, 3 g L⁻¹ KH₂PO₄ and 5 g L⁻¹ (NH₄)₂SO₄ pH 5.6) or Minimal Media without Nitrogen (MM-N) (25 g L⁻¹ glucose, 0.5 g L⁻¹ MgSO₄·7H₂O, 3 g L⁻¹ KH₂PO₄, 6.3 g L⁻¹ K₂SO₄ and 0.2 g L⁻¹ (NH₄)₂SO₄ pH 5.6).

2.1.2. Bacteria

*Escherichia coli* DH5α was used for plasmid amplification and grown in Lysogeny Broth (LB) (‘LB (Luria-Bertani) liquid medium’, 2006) where transformants were selected for on LB supplemented with kanamycin (50 µg mL⁻¹) at 37 °C.

*Agrobacterium tumefaciens* strain GV3101 (Van Larebeke et al., 1974) was constitutively grown at 28 °C in LB with 100 µg mL⁻¹ rifampicin and 50 µg mL⁻¹ gentamycin with transformants selected for and maintained with an additional 50 µg mL⁻¹ kanamycin.

2.1.3. *Saccharomyces cerevisiae*

*S. cerevisiae* BY4742 (Y10000; S288C isogenic yeast strain MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0) (Brachmann et al., 1998; Winston et al., 1995) was used to perform in yeast plasmid assembly. It was ordinarily grown in YPD at 30 °C; transformants were autotrophically selected for on 6.9 g L⁻¹ YNB supplemented with 770 mg L⁻¹ CSM without uracil (Formedium, Hunstanton).
2.2. Bioinformatics
2.2.1. *R. toruloides* homologous gene identification

*R. toruloides* homologues of desired *S. cerevisiae* S288C genes identified in the ‘Saccharomyces Genome Database’ (Cherry et al., 2012; Mortimer and Johnston, 1986) were identified using BLASTN searches against the annotated *R. toruloides* NP11 genome (assembly accession GCA_000320785.2) in Ensembl Fungi (Zerbino et al., 2018; Zhu et al., 2012). 1200 bp upstream of the identified *R. toruloides* homologues were then used as the respective gene promoters.

2.2.2. Codon Optimisation

Codon optimisation was performed using Codon Optimization OnLine (COOL) (Chin et al., 2014) where codon-usage from genome-wide usage patterns was derived from *R. toruloides* strain IFO 0559 (accession GCA_000988805.1). Putative 5’ splice sites (A Johns pers. comm.) located within designed constructs were manually removed and alternative codons utilised.

2.3. Plasmid construction

2.3.1. *R. toruloides* genomic DNA extraction

DNA was extracted from *R. toruloides* using an adapted phenol: chloroform method from Sambrook and Russell (2001) for yeast colonies on plates, adapted where MP Biomedicals Matrix A lysing beads (Santa Ana, CA) were used in a MP Biomedicals FastPrep-24 (Santa Ana, CA) instead of vortexing. DNA was recovered by ethanol precipitation and resuspended in TE (pH 7.6) as described by Sambrook and Russell (2001).

2.3.2. Amplification of DNA

Fragments of DNA used for cloning were amplified through PCR using Q5 DNA polymerase (NEB, Ipswich MA) and purified using a Thermo Fisher Scientific GeneJet PCR Purification Kit (Waltham, MA); during colony PCRs to verify presence of DNA within assembled plasmids PCR was achieved using Taq Polymerase in Standard Buffer (NEB, Ipswich MA); in both cases oligonucleotides were synthesised by Eurofins (Ebersberg).

2.3.3. Restriction digestion

Restriction digests were performed using restriction endonucleases purchased from either NEB (NEB, Ipswich) or Thermo Fisher Scientific (Waltham, MA) as
denoted in Table 2. Cloning fragments were then purified using a Thermo Fisher Scientific GeneJet PCR Purification Kit (Waltham, MA).

### Table 2 | Restriction endonucleases used

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Manufacturer</th>
</tr>
</thead>
<tbody>
<tr>
<td>BamHI Fast Digest</td>
<td>Thermo Fisher Scientific</td>
</tr>
<tr>
<td>PmlI-HF</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>SpeI-HF</td>
<td>New England Biolabs</td>
</tr>
<tr>
<td>AflII-HF</td>
<td>New England Biolabs</td>
</tr>
</tbody>
</table>

2.3.4. Plasmid recovery from micro-organisms
Plasmid constructs were isolated from *E. coli* using Thermo Fisher Scientific GeneJet plasmid miniprep kit, following the manufacturer’s protocol (Waltham, MA). *S. cerevisiae* plasmids were isolated and purified through Zymolase assisted alkaline lysis using Zymoprep Yeast Plasmid Miniprep II kit following the manufacturer’s protocol for single-picked colony plasmid extraction (Irvine, CA).

2.3.5. DNA gel electrophoresis
All DNA fragments were visualised in 2 % agarose gels (Melford) with Invitrogen Sybr Safe DNA stain (Thermo Fisher Scientific, Waltham, MA) unless otherwise specified. DNA bands were separated using electrophoresis at 100 V for approximately 60 minutes, until desired gel bands could be resolved. Gels were visualised through UV transillumination.

2.3.6. Bacterial transformation
*E. coli* DH5α was transformed using a modified protocol from Glover (1985). Competent cells were prepared by pelleting OD_{600} 0.4 mid-log *E. coli* before: washing in TF-1 (7.4 g L⁻¹ KCl, 2.95 g L⁻¹ CH₃COOH 1.5 g L⁻¹ CaCl₂.2H₂O, 150 g L⁻¹ glycerol, 0.2 M MnCl₂); cooling for 15 minutes on ice; resuspension in TF-2 (0.74 g L⁻¹ KCl, 11 g L⁻¹ CaCl₂.2H₂O, 150 g L⁻¹ glycerol, 0.01 M MOPS buffer pH 6.8) and snap freezing in liquid nitrogen. Competent cells were transformed with desired plasmid constructs by: thawing on ice; addition of 1-5 µg DNA; incubation on ice for 40 minutes; 2 minutes of heat shock at 42 °C; 5 minutes of incubation on ice; 60 minutes of recovery in 37 °C LB shaken at 200
RPM; plating on 1.5 % w/vol LB agar with 50 µg mL⁻¹ kanamycin; and incubation overnight at 37 °C.

*A. tumefaciens* GV3101 was prepared and transformed as described by Holsters *et al.*, (1978).

2.3.7. *S. cerevisiae* transformation

Yeast Recombination-Based Cloning (YRBC) was performed using a modified protocol from Kilaru and Steinberg, (2015) as DNA fragments with 40 bp end-regions homology, at a 2:1 fragment to base construct molar ratio, were co-transformed. Transformants were selected for on 3 % YNB agar supplemented with CSM without uracil.
### Table 3 | Primers used for DNA amplification to allow organelle visualisation in *R. toruloides*

<table>
<thead>
<tr>
<th>Primer</th>
<th>Function</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atg8_YRBC</td>
<td>F</td>
<td>YRBC</td>
</tr>
<tr>
<td>Atg8_YRBC</td>
<td>R</td>
<td>YRBC</td>
</tr>
<tr>
<td>Cal_YRBC</td>
<td>F</td>
<td>YRBC</td>
</tr>
<tr>
<td>Cal_YRBC</td>
<td>R</td>
<td>YRBC</td>
</tr>
<tr>
<td>Ldp1_YRBC</td>
<td>F</td>
<td>YRBC</td>
</tr>
<tr>
<td>Ldp1_YRBC</td>
<td>R</td>
<td>YRBC</td>
</tr>
<tr>
<td>Atg8 DS Check</td>
<td>F</td>
<td>Junction</td>
</tr>
<tr>
<td>Atg8 US Check</td>
<td>R</td>
<td>Junction</td>
</tr>
<tr>
<td>CMV35 DS Check</td>
<td>R</td>
<td>Junction</td>
</tr>
<tr>
<td>Calnexin DS Check</td>
<td>F</td>
<td>Junction</td>
</tr>
<tr>
<td>Calnexin US Check</td>
<td>R</td>
<td>Junction</td>
</tr>
<tr>
<td>Ldp1 US Check</td>
<td>R</td>
<td>Junction</td>
</tr>
<tr>
<td>Ldp1 DS Check</td>
<td>R</td>
<td>Junction</td>
</tr>
<tr>
<td>PGK US Check</td>
<td>F</td>
<td>Junction</td>
</tr>
<tr>
<td>EGFP US Check</td>
<td>F</td>
<td>Junction</td>
</tr>
<tr>
<td>EGFP DS Check</td>
<td>R</td>
<td>Junction</td>
</tr>
</tbody>
</table>

Lowercase orange italic letters signify homologous recombination regions, uppercase italic blue letters represent coding sequences for 6x glycine ‘linker’ sequences, and black uppercase lettering depicts specific bases for PCR amplification. All primers are written from 5’ to 3’. 

35
<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>01-FL</td>
<td>atcatcgcaagaccggcaacaggattcaatCTGCAACCTCCTCGCTGCTCG</td>
</tr>
<tr>
<td>01-IG</td>
<td>atcatcgcaagaccggcaacaggattcaatCCTGCAAGTCACCTCTGTTGCTGCT</td>
</tr>
<tr>
<td>01</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGGTGGGCTGCACGGGGGAGG</td>
</tr>
<tr>
<td>02-FL</td>
<td>atcatcgcaagaccggcaacaggattcaatGTCGCAGCTCCTGACCCGC</td>
</tr>
<tr>
<td>02-IG</td>
<td>atcatcgcaagaccggcaacaggattcaatGATGGCGAGTCTCGTTGAGAGA</td>
</tr>
<tr>
<td>02</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGTGTCGACGGTCAGTC</td>
</tr>
<tr>
<td>03-NC</td>
<td>atcatcgcaagaccggcaacaggattcaatCTGAGCAGTTCGGCAGTGA</td>
</tr>
<tr>
<td>03</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGGCGTCCCTGGAGAGTAGCC</td>
</tr>
<tr>
<td>04-FL</td>
<td>atcatcgcaagaccggcaacaggattcaatTATGCACGCTGAAAACGGCTCG</td>
</tr>
<tr>
<td>04-IG</td>
<td>atcatcgcaagaccggcaacaggattcaatTCGCTCCCGCTGCGAGTC</td>
</tr>
<tr>
<td>04</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGTGTCGACGGTCAGTC</td>
</tr>
<tr>
<td>05-FL</td>
<td>atcatcgcaagaccggcaacaggattcaatCAGACCCGTTGCCGACCTC</td>
</tr>
<tr>
<td>05-IG</td>
<td>atcatcgcaagaccggcaacaggattcaatGCCCACCGACGAACAGCTCC</td>
</tr>
<tr>
<td>05</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGTGTCGACGGTCAGTC</td>
</tr>
<tr>
<td>06-FL</td>
<td>atcatcgcaagaccggcaacaggattcaatATCTTGCCCGCTTTGCG</td>
</tr>
<tr>
<td>06-IG</td>
<td>atcatcgcaagaccggcaacaggattcaatCGGGGCGGACAGACGCCAG</td>
</tr>
<tr>
<td>06</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGTGTCGACGGTCAGTC</td>
</tr>
<tr>
<td>07-FL</td>
<td>atcatcgcaagaccggcaacaggattcaatGGGTGAGGTGCTGTTAGTGA</td>
</tr>
<tr>
<td>07-IG</td>
<td>atcatcgcaagaccggcaacaggattcaatTTGCCCATCCTTGCAACTCACA</td>
</tr>
<tr>
<td>07</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGTGTCGACGGTCAGTC</td>
</tr>
<tr>
<td>08-FL</td>
<td>atcatcgcaagaccggcaacaggattcaatGGGTGAGGTGCTGTTAGTGA</td>
</tr>
<tr>
<td>08-IG</td>
<td>atcatcgcaagaccggcaacaggattcaatTTGCCCATCCTTGCAACTCACA</td>
</tr>
<tr>
<td>08</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGTGTCGACGGTCAGTC</td>
</tr>
<tr>
<td>09-FL</td>
<td>atcatcgcaagaccggcaacaggattcaatGTGCCTTTCTCTGCGTCTGTCG</td>
</tr>
<tr>
<td>09-IG</td>
<td>atcatcgcaagaccggcaacaggattcaatGACAGGCGGCTGGCGCAAG</td>
</tr>
<tr>
<td>09</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGTGTCGACGGTCAGTC</td>
</tr>
<tr>
<td>10-FL</td>
<td>atcatcgcaagaccggcaacaggattcaatCTCAGGTCACCTCGGTCAGG</td>
</tr>
<tr>
<td>10-IG</td>
<td>atcatcgcaagaccggcaacaggattcaatGACAGGCGGCTGGCGCAAG</td>
</tr>
<tr>
<td>10</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGTGTCGACGGTCAGTC</td>
</tr>
<tr>
<td>11-IG</td>
<td>atcatcgcaagaccggcaacaggattcaatGTGCCTTTCTCTGCGTCTGTCG</td>
</tr>
<tr>
<td>11-FL</td>
<td>atcatcgcaagaccggcaacaggattcaatGACAGGCGGCTGGCGCAAG</td>
</tr>
<tr>
<td>11</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGTGTCGACGGTCAGTC</td>
</tr>
<tr>
<td>12-FL</td>
<td>atcatcgcaagaccggcaacaggattcaatTGGCCCATCCTTGGAACCTACA</td>
</tr>
<tr>
<td>12-IG</td>
<td>atcatcgcaagaccggcaacaggattcaatGACAGGCGGCTGGCGCAAG</td>
</tr>
<tr>
<td>12</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGTGTCGACGGTCAGTC</td>
</tr>
<tr>
<td>13-FL</td>
<td>atcatcgcaagaccggcaacaggattcaatAAAAAGAAGGGCAGGTCAGG</td>
</tr>
<tr>
<td>13-IG</td>
<td>atcatcgcaagaccggcaacaggattcaatGACAGGCGGCTGGCGCAAG</td>
</tr>
<tr>
<td>13</td>
<td>ttcctgcccttcctcaggacccgctccgcctcctcgtcGTGTCGACGGTCAGTC</td>
</tr>
</tbody>
</table>

*R. toruloides* HSPP primers (‘FL’ is Fixed Length 1200 bp, ‘IG’ is Intergenic’) and then directionality of the primer (‘F’ is ‘Forwards’, ‘R’ is ‘Reverse’). Primer sequences show homology regions to the base plasmid pEGFP-Rt-YR-Hyg in lowercase coloured italics and uppercase lettering depicts specific bases for PCR amplification.
2.4. Transformation of *R. toruloides*

*R. toruloides* transformation was informed by methods described by Johns *et al.* (2016) for transformation of *R. toruloides* and a protocol for *Zymoseptoria tritici* transformation (Nic Hemstetter pers. comm.), giving a transformation efficiency at least x200 above that described by Johns, Love and Aves (2016). *A. tumefaciens* containing the desired transformation cassette was grown to stationary phase in LB supplemented with *A. tumefaciens* virulence plasmid maintenance antibiotics rifampicin (100 μg mL⁻¹) and gentamycin (50 μg mL⁻¹) and plasmid selection antibiotic kanamycin (50 μg mL⁻¹) at 28 °C for 36 hours. This was then diluted to an OD₆₀₀ of 0.10 in induction medium (10 mM K₂HPO₄, 10 mM KHzPO₄, 2.5 mM NaCl, 2 mM MgSO₄, 0.7 mM CaCl₂, 10 μM FeSO₄, 4 mM (NH₄)SO₄, 10 mM glucose, 40 mM 2-(N-morpholino) ethanesulphonic acid, 0.5 % w/vol glycerol adjusted to pH 5.6 with NaOH or phosphoric acid), supplemented with 200 μM acetylsyrongone, 50 μg mL⁻¹ rifampicin, 25 μg mL⁻¹ gentamycin and 25 μg mL⁻¹ kanamycin and incubated for approximately 6 hours until OD₆₀₀ = 0.2. 500 μL *A. tumefaciens* and 500 μL mid-to-late log phase *R. toruloides* were mixed and spread over 325 P Clear Cellulose Film Discs (A. A. Packaging Limited, Lancashire, UK), placed on induction medium (as described previously, with 2 % agar) and incubated at 28 °C for 36 hours, until a pink colour began to develop. These cellulose membranes are then transferred to a YPD agar, supplemented with 150 μg mL⁻¹ cefotaxime and 50 μg mL⁻¹ hygromycin, incubated for 36-48 hours, and single colonies were restreaked to fresh YPD containing selective antibiotics and incubated for a further 36-48 hours for colonies to form.

2.5. CARS/SRS

Simultaneous Coherent Anti-Stokes Raman Spectroscopy (CARS) and Spontaneous Raman Spectroscopy (SRS) were performed exactly as described by Wang, *et al.*, (2018), except with a tuned Raman active mode of 2845 cm⁻¹.

2.6. Confocal microscopy

*R. toruloides* cells were cultured in their requisite media before immobilisation by 1:9 dilution in 15 % low melting temperature agarose (Melford, Suffolk, UK). Fluorophores were excited using a 488 nm OPSL 488 laser (Coherent Inc., Santa Clara, CA) and their fluorescence emission recorded using GFP channel filter (HyD 500-535 nm); autofluorescence channel setting (HyD 650-700 nm); and a brightfield image (PMT) in
a Leica TCS SP8 confocal laser scanning microscope (Wetzlar, Germany). The Leica was equipped with an HC PL APO 100x/1.44 oil objective at a scan speed of 200 Hz.

2.7. Quantification of lipid droplets in confocal fluorescence images

Side-profile *R. toruloides* cells were manually selected and their images recorded as described in 2.6.

Volumes of cells were ascertained using a custom ImageJ macro utilising plugins available in Fiji and a custom Python script (Oliphant, 2010; Schindelin *et al.*, 2012). Briefly, as described in Figure 4, the cell edge was selected using an appropriate brightfield slice using 4-way connectedness of similar pixel values; this cellular outline was then fitted to an ellipse to find major and minor axis angles and lengths and the original cell outline rotated about the angle of the major axis. The x and y co-ordinates for this rotated cell outline were then passed to a custom Python script to find maximum and minimum values for each y value along the x axis (to remove small artefacts in the area where connectedness had introduced invaginations of the cell profile), and the distance between the two termed the ‘slice height’. The volume for each x ‘slice’ was determined by multiplying each ‘slice height’ by the pixel height, pixel width and pi; the series of summed slices, projected about the x-axis, is equal to the calculated cell volume.

Volumes of lipid were calculated again using Fiji (Schindelin *et al.*, 2012). Briefly, the GFP channel was despeckled to reduce noise and Gaussian blurred. The resultant image was auto-thresholded using the Renyi entropy algorithm (Narayana *et al.*, 2012) before volumes of each fluorescent body were calculated using the plugin ‘3D Objected Counter’ (Bolte and Cordelières, 2006).
Figure 4| Ascertaining cell volume using a brightfield image. (A) A cell-outline was found from the middle slice of a side-profile cell (dark orange) and an ellipse fitted to this shape (green). The cell side-profile was then rotated to fit an x-axis aligned profile and invaginations of the cell outline (exaggerated in figure) accounted for and corrected, as in (B). Each X pixel ‘slice’ was then projected about 360° to find volumes for each slice before each x-slice was summed.

2.8. Quantification of yeast growth rate and lipid droplet responses to different media

*R. toruloides* Ldp1-GFP was precultured to stationary phase in YNB + CSM with 20 g L⁻¹ for 72 hours before: a single wash and dilution to OD₆₀₀ 0.2 in Minimal Media without Nitrogen (0 g L⁻¹); and then diluted 1:99 in Minimal Media without Nitrogen supplemented with 0, 5, 10, 15, 20, or 25 g L⁻¹ glucose; all incubations occurred at 30 °C.

1000 μL cultures were then loaded into a 4titude Vision 24 Plate and incubated at 500 RPM double orbital shaking, 30 °C for 72 hours in a BMG ClarioStar plate reader (Buckinghamshire, UK). Absorbance and GFP were measured every 10 minutes; for each well both values were determined using 20 orbital flashes well⁻¹; average well absorbance values were path-length corrected to 1 cm and GFP was measured using 428+14 nm excitation and 530+30 nm notch filters.

After 72 hours, cells were visualised in an Amnis ImageStream X Mark II Imaging Flow Cytometer (Brooks Life Science, Chelmsford, CA) through brightfield and detected fluorescence at 488 nm. Side profile cells were found by first gating a subpopulation of sharp-focused cells using brightfield intensity gradient (gradient relative mean squared 53.96 - 84.35), then gating a brightfield Aspect Ratio: Area slice from approximately 22:0.96 - 60.3:0.67, a circularity score between 5.4 - 8.4, a gating of
brightfield Major Axis of 3-10 before finally removing cells with no GFP fluorescence detected. Values for side-profile GFP and Brightfield areas were determined using the ‘Threshold’ and ‘Inspire’ mask functions.

2.9. Measuring single-cell HSPP induction

*R. toruloides* transformed with EGFP controlled HSPP promoters were grown in YNB + CSM (20 g L⁻¹ glucose) at 28 °C until late-log stage before dilution to OD₆₀₀ 0.2 in pre-warmed 28 °C YNM + CSM (20 g L⁻¹ glucose) and shaken for 60 minutes at 28 °C. This was followed by either heat shock to 37 °C or remaining at 28 °C for 60 minutes before further incubation at 28 °C for 5 hours. All incubation steps were at 200 RPM.

Samples were then pelleted and resuspended in F1 (4 g L⁻¹ paraformaldehyde, 3.4 g sucrose) and incubated at 21 °C for 15 minutes. These were then washed in F2 (1.2 M sorbitol, 0.1 M potassium phosphate pH 7.5) before resuspension in F2. Fluorescence of each cell was then quantified using an Amnis ImageStream X Mark II Imaging Flow Cytometer (Brooks Life Science, Chelmsford, CA) using settings for side-profile cell determination described in section 2.8; the fluorescence value determined is the median pixel within the ‘Inspire’ mask.

2.10. Measuring HSPP induction kinetics

*R. toruloides* transformed with EGFP controlled HSPP promoters were grown in YNM + CSM (20 g L⁻¹ glucose) at 25 °C until late-log stage before dilution to OD₆₀₀ 0.2 in pre-warmed 25 °C YNM + CSM (20 g L⁻¹ glucose), loaded into 4titude 28-well plate and shaken for 60 min at 28 °C. This was followed by either heat shock to 37 °C or remaining at 28 °C for 120 min before further incubation at 28 °C for 4 hours. All incubation steps were at 200 RPM.
3. Results & Discussion: Organelle visualisation in \textit{R. toruloides}

As outlined in section 1.1, \textit{R. toruloides} is an oleaginous yeast with potential commercial uses as a more environmentally responsible method for producing carotenoids, biofuels, and other lipid-derived compounds. Work has been done to improve lipid production above the levels in wild-type yeast, which are capable of accumulation of lipid droplets to over 76 \% dry biomass (Li, Q. \textit{et al.}, 2008; LI \textit{et al.}, 2006), through overexpression of native acetyl-CoA carboxylase and diacylglycerol acyltransferase (Zhang, Skerker, \textit{et al.}, 2016). However, \textit{R. toruloides} is underdeveloped as a biotechnological chassis and requires further investigation to improve efficiency and rate of lipid production, control of gene expression, as well as improving ease of lipid/biofuel release for downstream processing. This chapter describes studies to enable the visualisation of aspects of lipid droplet dynamics and interactions in \textit{R. toruloides}. In the next chapter, tools will be sought for the improved control of gene expression for biotechnological exploitation of this yeast.

3.1. Construction of GFP-tagged protein expression plasmids

To develop a toolkit to enable live-cell visualisation and monitoring of the lipid droplet, ER, and autophagy complex, it was necessary to fluorophore epitope-tag proteins known to localise to these organelles: Ldp1 (Zhu \textit{et al.}, 2015), calnexin (Brandizzi \textit{et al.}, 2004; Runions \textit{et al.}, 2006), and Atg8 (Li, D. \textit{et al.}, 2015) respectively. Ldp1 is a protein which localises to the boundary of lipid droplets, and its tagging with a fluorophore enables this organelle to be directly visualised in living yeast/fungal cells (Zhu \textit{et al.}, 2015). Calnexin localises to the ER, the site of site of biogenesis of the lipid droplet and the principal site of TAG synthesis for lipid droplet formation. Tagging calnexin can enable the dynamics of lipid droplet formation to be monitored (Müller-Taubenberger \textit{et al.}, 2001). Atg8 protein is recruited to degrade lipid droplet lipids for FFA recycling through microautophagy (Hashemi and Goodman, 2015; Wang, C. W., 2016). Tagging of these three proteins should therefore enable the dynamics of \textit{R. toruloides} lipid droplets to be monitored via live-cell visualisation at all stages of their “life cycle”.

Successful studies on the Ldp1, calnexin and Atg8 proteins in other organisms have shown specificity for either -C or -N terminus tagging (Klionsky, 2011; Müller-Taubenberger \textit{et al.}, 2001; Zhu \textit{et al.}, 2015) for localised organelle-targeted fluorescence, and this information was used when designing tag constructs. A codon-
optimised eGFP tag was used which has been previously demonstrated to function efficiently in *R. toruloides* (Johns et al., 2016). For transformation of *R. toruloides*, the base plasmid pEGFP-Rt-YR-Hyg was used (Figure 5), which contains a codon-optimised eGFP gene under control of the native *R. toruloides* PGK1 promoter, plus elements for maintenance and selection in *S. cerevisiae*, *E. coli*, and *Agrobacterium tumefaciens* to enable transformation of *R. toruloides* by *A. tumefaciens*-mediated transformation (ATMT) (Johns et al., 2016; Zhang, Skerker, et al., 2016).

The strategy for constructing each of the GFP-tagged protein expression plasmids is given in Figure 6. This strategy utilised Yeast Recombination-Based Cloning (YRBC), an *in vivo* gene assembly method which utilises the highly efficient and accurate homologous recombination mechanism within the budding yeast *S. cerevisiae*; multiple linear DNA fragments with short (40 bp) overlapping sequences are transformed into *S. cerevisiae*, accurately joined to form a plasmid *in vivo*, and recovered via DNA preparation and transformation of *E. coli*. YRBC is particularly suitable for assembly of large and complex plasmids such as those required for *Agrobacterium tumefaciens*-mediated transformation (ATMT) of fungi (Johns et al., 2016; Kilaru and Steinberg, 2015).

Native *R. toruloides* CBS 14 strain sequences encoding the calnexin and Atg8 proteins were identified using reciprocal BLAST searches as specified in section 2.2.1. The *R. toruloides* Ldp1, calnexin and Atg8 genes were amplified with gene sequence-specific primers containing an additional 5′ 40 bp homology region to either: PGK1 promoter or 5′ EGFP with an additional codon-optimised 6x glycine residue linker sequence (*LDP1* and calnexin); or a 5′ 40 bp homology region to either: EGFP with an additional codon-optimised 6x glycine residue linker sequence or 5′ 40 bp homology sequence to CMV35S terminator as detailed in Table 3, Section 2. To ensure only the desired amplicon was produced in the PCR reaction, PCR products were size-fractionated by agarose gel electrophoresis to verify the fragment size: 1327 bp (*LDP1*), 2428 bp (calnexin), and 876 bp (*ATG8*) as shown in Figure 7A.

The base plasmid pEGFP-Rt-YR-Hyg (Figure 5) was cleaved with either *AflII* (*LDP1* and calnexin) or *PmlI* (*ATG8*), and the *R. toruloides* protein coding DNA amplicons were integrated adjacent to GFP using *Saccharomyces* YRBC, as shown in Figure 6.
This led to the construction of the *R. toruloides* tagging plasmids pLDP1-EGFP-Rt-YR-Hyg, pCalnexin-EGFP-Rt-YR-Hyg, and pEGFP-Atg8-Rt-YR-Hyg respectively.

After growth of putatively transformed *S. cerevisiae* BY4742 on selective media, yeast colony PCR was performed on resultant colonies to screen for desired plasmid constructs using the ‘check’ primers, flanking each YRBC junction as demonstrated in Figure 6 with a 5′ ‘upstream’ junction and 3′ ‘downstream’ junction. Expected upstream amplicon sizes of 275 bp, 236 bp, and 343 bp and downstream amplicon sizes of 567 bp, 498 bp and 452 bp for the constructs containing *LDP1*, calnexin, and *ATG8* were found and are shown in Figure 7B. Plasmids from colonies with the correct junctions were then isolated, and these were then transformed into *E. coli* DH5α. Subsequent amplified constructed plasmid was then miniprepped from *E. coli* and verified by restriction-digestion with *Bam*HI; major expected bands of the constructs containing all contained common elements 14071 bp, 250 bp, and 21 bp with additional identifying bands of *LDP1* (1658 bp, 946 bp), calnexin (2759 bp, 946 bp), and *ATG8* (1753 bp, 403 bp) were all observed, as demonstrated on Figure 7C.
Figure 5] Map of base vector pEGFP-Rt-YR-Hyg for transformation of *R. toruloides*. This base plasmid as described by Johns et al. (2016) contains a T-DNA region between the left border (LB T-DNA) and right border (RB T-DNA) which becomes integrated into the *R. toruloides* genome, as well as sequences for propagation in *S. cerevisiae* (2µ ORI and URA3), and sequences for propagation in *E. coli* and *A. tumefaciens* (remainder of vector). The T-DNA region contains a codon-optimised eGFP gene regulated by the constitutive *R. toruloides* PGK1 promoter and the CMV35S terminator, and a codon-optimised hygromycin resistance marker, for selection in *R. toruloides*, regulated by the GPD1 promoter and *Neurospora crassa* β-tubulin (NcB Tub) terminator.
Yeast Recombination-Based Cloning (YRBC) strategy for construction and verification of plasmids for organelle visualisation in *R. toruloides*. Amplicons were generated by PCR containing an *R. toruloides* gene and flanking homology regions. GFP fusion constructs were then generated by homologous recombination with cleaved *R. toruloides* vector pEGFP-Rt-YR-Hyg (Figure 5), in *S. cerevisiae*. A. C-terminus GFP-tagged *LDP1* and calnexin amplicons generated through PCR of the respective *R. toruloides* genes using specific forward (F) primers, with an additional 5’ 40 bp homology region to the *PGK1* promoter, and reverse (R) primers, with a 5’ 40 bp homology region to eGFP, and a ‘linker’ sequence encoding 6x glycine residues. Amplicons were recombined via YRBC with cleaved pEGFP-Rt-YR-Hyg as illustrated. B. An N-terminus GFP-tagged *ATG8* amplicon was generated using the same strategy as described in A, but with the forward primer including a 5’ 40 bp eGFP homology sequence before a linker sequence, and the reverse primer containing a 5’ 40 bp CMV35S Terminator homology sequence. The amplicon is recombined via YRBC as illustrated. Relevant junction-flanking primers are also annotated (US ‘Upstream’/DS ‘Downstream’ Check) to show their approximate locations for later verification of YRBC constructs. Primer sequences are given in Table 1 (section 2.2). Diagrams are not to scale.
Figure 7 | Construction by YRBC of *R. toruloides* GFP-tagged protein expression plasmids for organelle visualisation. Marker lanes are labelled ‘M’ and are loaded with NEB Log2 Ladder. **A.** Separate gels containing genes amplified through PCR from extracted *R. toruloides* CBS14 genomic DNA using primers as described in Table 4 (Section 2) to include an additional 80 bp of YRBC-required homology region above genomic CDS length. *LDP1* was loaded into lane 1 of a 1 % gel with an expected band size of 1327 bp. Calnexin and *ATG8* were loaded into lanes 2-3 of 2 % gels with an expected band size of 2428 bp and 867 bp respectively. Track M contains NEB Log2 ladder. **B.** Yeast colony PCR of junctions between pEGFP-Rt-YR-Hyg and either the Ldp1 (left), calnexin (middle) or Atg8 (right) gene, using the strategies outlined in Figure 6. In all cases, the first odd numbered tracks of each gel contained PCR products using upstream junction primers and even numbered tracks contained the applicable downstream junction primers; lanes 3-4, 7-8, and 11-12 are negative PCR controls. *LDP1*, calnexin, and *ATG8* junctions had an expected band size of 275 bp, 236 bp, and 343 bp for their upstream junctions and 567 bp, 498 bp, and 452 bp for their downstream junctions respectively. Track M contains NEB Log2 ladder. **C.** Diagnostic *Bam*HI digests of YRBC-formed plasmids for pLdp1-EGFP-Rt-YR-Hyg (lanes 1-2, expected dominant bands of 14017 bp, 1658 bp, 946 bp and 250 bp), pCalnexin-EGFP-Rt-YR-Hyg (lanes 3-4, expected dominant bands of 14017 bp, 2759 bp, 946 bp and 250 bp), and pEGFP-Atg8-Rt-YR-Hyg (lanes 5-6, expected dominant bands of 14017 bp, 1753 bp, 403 bp, 250 bp) with undigested negative controls in the even numbered tracks.
3.2. Visualisation of lipid droplets in *R. toruloides*

Ldp1 (lipid droplet protein 1) is a structural protein identified in the lipid droplet proteome of *R. toruloides* (Zhu et al., 2015). An Ldp1-GFP construct expressed heterologously in *S. cerevisiae*, albeit with a different linker sequence, led to supersize lipid droplet formation (Zhu et al., 2015). The most abundant lipid droplet-associated protein identified in the proteome analysis of the *R. toruloides* is Ldp1 (Zhu et al., 2015). Sequence analysis of Ldp1 shows similarity to the perilipin family of proteins; the primary function of perilipins is the inhibition of lipase activity against lipid droplets by coating the organelle and physically restricting access (Sztalryd and Brasaemle, 2017; Zhu et al., 2012). Recent evidence has emerged to support additional functions of the perilipins: they both promote assembly of lipid droplets in the ER membrane and stabilise the phospholipid (PL) monolayer of nascent (<200 nm diameter) and pre-lipid droplets, preventing fusion with otherwise ‘unprotected’ lipid droplets as can be observed in *pet10Δ S. cerevisiae* perilipin mutants (Gao et al., 2017; Hashemi and Goodman, 2015).

Lipid droplet visualisation in live-cell images will be crucial for understanding lipid droplet biogenesis, lipid turnover, and lipid metabolic responses in oleaginous yeasts to external stresses and media, as well as the interactions of lipid droplets with other organelles. While conventional lipid stains, such as Nile Red, have been used to stain *R. toruloides* lipid droplets previously (Alexander Johns, pers. comm.), this stain is limited in its use for live-cell imaging of this yeast. This is primarily due to the extensive overlapping excitation and emission spectra of the stain and carotenoids of the yeast which preclude studies in stationary-phase cells, in addition to having to introduce the stain itself, which may induce artefacts into the sample or alter the lipid droplet structure. Due to the ability of *R. toruloides* Ldp1 to localise exclusively to lipid droplets in *S. cerevisiae*, this protein was chosen for a GFP fusion to attempt to visualise the lipid droplet in *R. toruloides* using the native *LDP1*-encoded perilipin.

*R. toruloides* was transformed through ATMT with the pLdp1-EGFP-Rt-YR-Hyg construct, using a modified protocol described in section 2.4 which generated a 200-fold greater number of transformants per plate than that reported by Johns et al., 2016. To demonstrate that Ldp1-GFP specifically targets the lipid droplet, confocal microscopy was used to excite and record GFP fluorescence (Figure 8). In this figure brightfield image signal is thresholded to highlight only peaks in
signal. The spherical GFP-coated structures found closely resemble the lipid droplet shape that we would expect to observe were lipid droplet to be enveloped in the fluorophore. However, to be able to conclusively determine lipid droplet/Ldp1-GFP targeting, verification is required with colocalisation of GFP signal to a known lipid droplet marker.

**Figure 8** Single slice confocal image of *R. toruloides* cells transformed with Ldp1-GFP. *R. toruloides* CBS14 transformed with pLdp1-GFP-Rt-YR-Hyg plasmid after growth in MM+N for 48 hours was visualised with GFP excited at 488 nm and Stokes shift fluorescence detected at 500-535 nm; signal shown here in green. The brightfield image was thresholded to show signal peaks associated with ‘halo’ artefacts in density-change.
To show colocalisation of Ldp1-GFP to lipid droplets, it was initially planned to use simultaneous Two-Photon-EpiFluorescence (TPEF), to excite the GFP fluorophore, and a coherent pump and Stokes laser to vibrationally excite the unlabelled CH$_2$ bond of acyl-fatty acids for spatial mapping of the compound using Stimulated Raman Loss (SRL). As $\omega_s - \omega_p = \Omega$, and the Raman active mode ($\Omega$) of CH$_2$ is 2845 cm$^{-1}$ (3514 nm), neither the tuneable pump beam (798 nm), fixed frequency-doubled 1032 nm Stokes beam, nor outputted antistokes beam of 650 nm are able to excite the 488 nm peak GFP fluorophore. However, although more convoluted, it is possible to use CARS/SRS to show colocalisation of Ldp1-GFP and lipid droplet. Using a standard SP8 confocal microscope, GFP excitation is shown to colocalise with brighter regions in a brightfield channel as light passes through a different-phase area as shown in Figure 8 (with GFP localised to the periphery of the circular sub-cellular structures). Correspondingly, in Figure 9, brightfield-channel highlighted regions in B are spatially co-localised with peak intensity regions in CARS/fluorescence channel in A and SRS channel in C, with a ‘Pearson Co-localisation Coefficient’ of 0.54 between brightfield and background-subtracted CARS. As such, this microscopy data suggests that Ldp1-GFP is able to precisely localise to the periphery of organelles containing the FA palmitic acid, found within the lipid droplet of *R. toruloides*.

**Figure 9** Composite images of CARS/SRS and Ldp1-GFP tagged *R. toruloides*. *R. toruloides* CBS14 was transformed with an Ldp1-GFP containing transformation cassette via ATMT. The transformants were cultured in Minimal Medium without Nitrogen for 48 h before visualisation on a CARS/SRS-capable microscope. **A.** Channel showing excitation with a CARS and fluorescence signal. **B.** Brightfield image of the yeast. **C.** SRS channel fluorescence. Scale bar = 15 µm
3.3. Development of a precise and a rapid method for quantification of lipid biogenesis

Having established that Ldp1-GFP localises to the periphery of lipid droplets, as suggested by Zhu et al., 2015, we then wanted to establish whether it would be possible to use this as a proof-of-principle method to calculate volumes of lipid droplet within each cell and, therefore, quantify rate of and efficiency of lipogenesis in different media. Both cell and lipid droplet volumes were determined, as described in section 2.7, where confocal stacks were taken of side-profile Ldp1-GFP transformed R. toruloides yeast before analysis in ImageJ. The relationship between fluorescence-microscopy images (left) and lipid droplets identified by the ImageJ macro (right), where individual lipid droplets are assumed to be separated if there is no GFP-signal connectedness between their surfaces, can be shown qualitatively in Figure 10A where a representative image is shown. The yeast cell wall was identified in the brightfield channel and is outlined in orange; the cell volume for this yeast was calculated by projecting this orange-outlined profile 360° around the major axis. Figure 10B shows the relationship between cell volumes and lipid volumes determined through this method when grown in either lipid accumulating MM-N (minimal media without nitrogen; red) or MM+N (minimal media with nitrogen; green). As nitrogen starvation inhibits both amino acid and nucleic acid synthesis, in common with other yeasts, it is assumed that this will halt the cell cycle during G1 with a consequence of similarly sized cells with an increased TAG content (and lipid droplet volume) in nitrogen-limited medium. However, no statistical difference could be identified for neither cell volumes nor lipid volumes between the two growth conditions using a Student’s t test. When cultured in MM-N, cell volume was an average of 78 µm³ ±51 µm³ and lipid droplet volumes of 18 µm³ ±12 µm³ (n=16); when cultured in MM+N cell volume was an average of 122 µm³ ±78 µm³ and lipid droplet volumes of 27 µm³ ±20 µm³ (n=7). Only a small number of samples were able to be analysed on the confocal microscope due to limited microscope time being available.
Figure 10| Comparison between raw confocal measurements of GFP-labelled Ldp1 organelles and the corresponding quantification of fluorescent intercellular compartments. A. Representative image of Maximum Projection areas identified and the volumes calculated in ImageJ. Green areas represent areas where voxels were included in the totalling of the lipid sum and white numbers labelling each lipid droplet and showing the centres of mass from of each individual vesicle within the image as calculated in section 2.7. Volumes of both whole cells and Ldp1-GFP-tagged subcellular structures were quantified as described in the 2.7. B. Quantitative calculated cell and lipid volumes using GFP-Ldp1, as described above, where 16 cells grown in MM-N for 48 hours (red markers) and 7 cells cultured in MM+N for 48 hours (green markers) are recorded.

3.4. Quantification of lipid biogenesis under glucose starvation conditions

To investigate whether available glucose concentration in the medium affects the rate of lipid droplet biosynthesis in R. toruloides, stationary phase GFP-Ldp1-transformed R. toruloides in MM+N was diluted into MM-N OD$_{600}$ of 0.02, with differing concentrations of glucose from 0 g L$^{-1}$ to 25 g L$^{-1}$, and the OD$_{600}$ recorded every 15 minutes for the next 72 hours. As shown in Figure 11 it is unsurprising that yeast cell density, determined by recording light scattering, began to rise earliest in the culture with the greatest concentration of reducible carbon, approximately 14 hours earlier than yeast cultured in the absence of glucose. As shown in Figure 11, the maximum rate of cell proliferation exhibited in MM-N is 1.2x10$^7$, 1.0x10$^7$, 1.1x10$^7$, 1.0x10$^7$, 7.9x10$^6$ and 5.8x10$^6$ cells mL$^{-1}$ h$^{-1}$ for 0, 5, 10, 15, 20, and 25 g L$^{-1}$ glucose respectively, where an OD$_{600}$ of 1 is approximately 5x10$^8$ cells mL$^{-1}$.
In contrast to the method for determining absolute cell-by-cell lipid droplet content using confocal microscopy above, where low sample numbers are likely to have prevented identifying statistical differences in lipid content biogenesis, we were curious to find whether it would be possible to determine any glucose-dependent differences in relative lipid droplet size with a cheaper high-throughput method. To achieve this, *R. toruloides* was inoculated as described in the glucose-dependent cell density experiment above (and section 2.8) before analysing >44 000 side-profile cells for each sample type through a flow cytometer; cell area was determined by using a root mean squared (RMS) gradient to identify cell edges, and lipid area was determined by summing known-dimension pixels fluorescing under 488 nm excitation. Figure 12 shows the plot of cell area and against lipid area from this flow cell analysis; although less useful than plotting lipid volumes for each culturing condition the resolving power of the Amnis flow cytometer was unable to distinguish each intracellular lipid droplet from one-another. Morphologically, a clustering of lipid areas and cell areas can be observed with 5-15 g L\(^{-1}\) initial glucose feeds, although as the glucose concentrations increase the lipid and cell areas appear to both decrease. While it may now be tempting to infer a relationship between the media and the cell lipid assimilation profile, this does not take into account the different growth stages of each of the yeast cultures at the point of data recording. While the ‘cluster’ of 5-15 g L\(^{-1}\) glucose are all still at logarithmic growth (or just entering diauxic shift), cells in 20 g L\(^{-1}\) glucose achieved stationary phase four hours before recording and cells in 25 g L\(^{-1}\) glucose had reached quiescence more than 20 hours previously. As yeasts begin to degrade lipid droplets in response to glucose limitation, to liberate and metabolise FFA, it is not surprising that cells which have apparently depleted a nutrient in their environment (and therefore reduced/stopped cell proliferation) have smaller lipid droplets.
Figure 11 | Rates of growth of GFP-Ldp1-transformed *R. toruloides* CBS14 with different glucose concentrations. Culture densities of *R. toruloides*, measured using OD$_{600}$, with different MM-N culture media with 0-25 g L$^{-1}$ glucose supplement, as indicated (right-hand side).
Median area of fluorescent regions within Ldp1-GFP-transformed \textit{R. toruloides}. Yeast were cultured in YNB+CSM for 36 hours before paraformaldehyde fixation and sample loading into an Amnis ImageStream Mark II X. >44 000 side-profile cells were measured and their areas were determined after excitation with a 488 nm laser. Median areas are plotted; error bars represent Median Absolute Deviations.

### 3.5. Endoplasmic reticulum visualisation in \textit{R. toruloides}

As described in section 1.1, the endoplasmic reticulum and lipid droplet organelles are intrinsically associated; the ER being the terminal site of the neutral lipid biogenesis pathway, the site of maintenance of the lipid droplet, the organelle which buffers against ER stresses, and the origin of many lipid droplet trafficked proteins. Throughout the lifecycle of the ER the ER and lipid droplet remain almost constantly in contact potentially through a ‘lipid bridge’ (Schuldiner and Bohnert, 2017). Markers for the transition zone cannot be resolved from one another through microscopy (Soni \textit{et al.}, 2009). As such, it is crucial to developing our understanding of the oleaginous yeast, as well as organelle dynamics and interactions, that we are able to live-cell visualise the ER organelle. To achieve this, the ER calcium storage protein/chaperone ‘calnexin’ was fused to GFP with a 6x codon-optimised glycine residue linker through genetic engineering and the construct was transformed into \textit{R. toruloides} as described in sections 2.3 and 3.1.
Without a counterstain which also targets the ER, it is unfortunately not possible to confirm calnexin localisation to the ER. However, as shown in Figure 13, GFP-fluorescent ‘tubules’ appear to have developed along the midline of the cell, with a secondary colocalisation to the two different-density organelles identified by their ‘halos’ in the brightfield channel – presumably the nucleus or lipid droplets. These findings support the notion that GFP-calnexin has localised to the ER, as this is where we would expect to find them, but further investigation is needed to verify this.

Figure 13| Z-stack montage showing localisation of calnexin-GFP in an example R. toruloides cell. Confocal slices showing areas of excitation at 488 nm (green) with an autofluorescence channel (650-700 nm) subtracted against a brightfield image of the yeast (grey) which was edited to highlight the refractive ‘halo’ artefact at the cell boundary. Cells were cultured in Minimal Medium without Nitrogen for 48 h before visualisation.

3.6. Autophagy visualisation and induction in R. toruloides
Autophagy (“self-eating”) is a eukaryotic cellular process required in the turnover and recycling of organelles and cytoplasm implicated in the (un)selective targeted degradation of many organelles. Whereas in yeast, lipolysis occurs either through
embedded lipid droplet lipases and sterol ester hydrolases, or translocation of the entire lipid droplet to vacuoles, mammalian cells are able to undergo macroautophagy, where the entire lipid droplet organelle is enveloped by an LC-3 (Atg8 homologue)-recruited double membrane. Development of this heterologous mammalian-like macroautophagy may be of marked interest in *R. toruloides*. Targeted autophagy research is still in its infancy, but lysosomal-associated membrane protein 2 (LAMP2) has been demonstrated to translocate to lysosomal fractions, concurrent with an increase in LC3 fraction-associated increase (Cahová, 2012). Targeting of a double-membrane organelle to the lipid droplet, where lipid reservoirs account for over 70% w/w, may be useful for inducible secretion of the mature LD-containing organelle; as the only organelle capable of *de novo* double-membrane engulfment, this autophagy complex may facilitate an energetically cheap method for releasing accumulated lipid into the surrounding medium. As outlined in section 1, in a pilot study over half of the cost of the biodiesel product was derived from energy costs in isolating the lipid. Targeted release may make this second-generation biofuel commercially competitive with fossil fuels, and therefore reduce GHG emissions.

As described in 4.2, an Atg8-GFP construct was made and transformed into *R. toruloides*. Having cultured the yeast in MM-N for 48 hours, a sample of cells from each of the biological replicates were visualised through confocal microscopy; a representative image of this microscopy is shown in Figure 14. While, excitingly, Figure 14 does appear to show puncta with an association to the region of budding and a second region of higher GFP co-localisation to the artefact on brightfield imaging, this could be due to chance or accumulation of the fluorescent protein as it precipitates from solution, or accumulates in inclusion bodies. As such, from our data, we cannot yet confirm that Atg8-GFP is associated with the autophagy complex, a necessity for visualising the targeted secretion of lipid droplet mentioned above. Further studies are required to verify the colocalisation of Atg8-GFP to the autophagy complex, either through a counterstaining experiment such as immunofluorescence staining against major autophagy proteins or identification of more fluorescent puncta as the yeast are exposed to autophagy-inducing conditions such as glucose starvation.
Figure 14 | Example image showing localisation of GFP-Atg8 in *R. toruloides*. A single confocal slice is showing areas of excitation at 488 nm (green) with an autofluorescence channel (650-700 nm) subtracted against a brightfield image of the yeast (grey) which was edited to highlight the refractive ‘halo’ artefact at the cell boundary. Cells were cultured in Minimal Medium without Nitrogen for 48 h before visualisation.
4. Results & Discussion: Heat Shock Protein Promoter Induction

As outlined in section 1.2, while there has been a rapid development of genetic tools in *R. toruloides*, all of the current eight inducible promoters (*NAR1, ICL1, MET16, CTR3, DAO1, THI4, THI5 and CTR31*) in the yeast rely on nutrient-limited media (Park *et al.*, 2018). This restricts their use in a commercial setting where feed is not tightly controllable or, if feed is tightly controllable, then metabolic remodelling to accommodate the limited nutrients may have unintended physiological effects on the yeast and reduce yield of the desired specialist products. As such, a short-term environmental assault, which would induce the Environmental Stress Response (ESR) through upregulation of specific genes’ transcription in the yeast, was decided as a candidate for inducible promoter investigation. Of the many assaults where global transcription has been monitored as part of an ESR (Causton *et al.*, 2001; Gasch, A. P. *et al.*, 2000), only ionising radiation or temperature shock do not involve any depletion, or changes, of composition of the media. It was therefore decided to use heat shock as the inducible ESR of choice as ionising radiation, where genes related to DNA repair are upregulated, may lead to unintended mutant variants being generated.

4.1. Promoter design and synthesis

*S. cerevisiae* genes were identified which were either up-regulated primarily in response to a heat-shock, as determined through microarray analysis, or which responded to a general ESR where the gene transcripts were up-regulated in response to many different types of environmental responses (Causton *et al.*, 2001; Gasch, A. P. *et al.*, 2000). Homologues of these selected ‘heat shock proteins’ (HSP) genes were then identified in *R. toruloides* (Table 5) and their promoters (heat shock protein promoters; HSPPs) chosen as candidates for inducible ESR-driven GFP-reporter synthesis. Forward-genetic studies by Liu, Y. *et al.*, 2018, where carotenoid biosynthesis was disrupted by random integrational mutagenesis through ATMT, showed this technique predominantly resulted in random single-copy chromosomal-integration of transfer-DNA. Previous TAIL-PCR experiments by Stephen Aves and Alexander Johns (pers. comm.) had not found any statistical differences in levels of gene expression dependent upon the site of DNA integration and were, therefore, not determined in this study.
It was also decided to investigate whether sequences of the immediately-upstream protein-coding gene contributed to HSPP induction, perhaps through transcription-factor binding in *R. toruloides* outside the intergenic region of the HSPP; the genetic schematics of the four inducible promoters previously characterised by Johns *et al.*, 2016 are shown in Figure 15. In the case of these promoters, *MET16*, *CTR3*, and *NAR1* all contained their crucial promoter regulatory elements within their intergenic regions. However, two distinct regulatory regions were identified in *ICL1* from -100 to -200, and from -200 to -400 bp. As *ICL1* intergenic region is only 218 bp, it is possible that a regulatory element is therefore found within the upstream gene; it was therefore decided to investigate whether there is any induction difference between ‘fixed-length’ (FL) 1200 bp or variable-length ‘intergenic’ (IG) promoters. This would contribute to our ability to more rapidly characterise *R. toruloides* promoters in the future as, if expression of the promoters is consistently the same between FL and IG promoters, then the number of future nested deletions of promoters (to find minimally-sized active promoters) can be reduced beyond an arbitrary ‘upstream’ length to just within the intergenic region. To test this, a yeast recombination-based cloning strategy was employed as described in Figure 16A. Briefly, promoter sequences of either fixed-length (FL) 1200 bp or a variable length (until the immediately upstream gene’s protein coding region; intergenic (IG) length) were identified from the annotated *R. toruloides* genome (Zhu *et al.*, 2012) and primers designed with sequence specificity to either: the FL sequence, with an additional 40 bp 5’ homology sequence to the base plasmid pEGFP-Rt-YR-Hyg immediately upstream of the PGK promoter, or with sequence specificity to the reverse strand 3’ end of the HSPP gene of interest with an additional 5’ 40 bp homology to EGFP. These promoters were amplified by PCR from *R. toruloides* genomic DNA and size-verified, as shown in Figure 17.
Figure 15| Schematic of inducible promoters previously characterised in \textit{R. toruloides} CBS14. A schematic diagram (not to scale) showing directionality and distance of upstream (US) protein-coding regions in relation to previously characterised methionine-inducible \textit{MET16}, copper-inducible \textit{CTR3}, nitrogen-inducible \textit{NAR1}, and acetate-inducible \textit{ICL1} genes in \textit{R. toruloides} CBS14; distances to the start or stop codons are shown in base-pairs.
Table 5 | *R. toruloides* heat shock protein promoters

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>S. cerevisiae Standard Name</th>
<th>Function</th>
<th>Full Name</th>
<th>R. toruloides Locus &amp; Direction</th>
<th>Promoter transcriptional interference</th>
<th>Promoter intergenic region lengths (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RHTO_00033</td>
<td>PGK1</td>
<td>Glycolysis</td>
<td>Phosphoglycerate kinase</td>
<td>Scaffold1: 74,979 - 76,478 : 1</td>
<td>T</td>
<td>373</td>
</tr>
<tr>
<td>RHTO_00032</td>
<td>ENO2</td>
<td>Glycolysis</td>
<td>Enolase II</td>
<td>Scaffold1: 783,641 - 786,135 : -1</td>
<td>D</td>
<td>873</td>
</tr>
<tr>
<td>RHTO_00384</td>
<td>KAR2</td>
<td>Chaperone</td>
<td>Hsp70 family ATPase</td>
<td>Scaffold1: 935,362 - 936,861 : -1</td>
<td>D</td>
<td>907</td>
</tr>
<tr>
<td>RHTO_01329</td>
<td>TPI1</td>
<td>Glycolysis</td>
<td>Triosephosphate isomerase</td>
<td>Scaffold13: 29,368 - 30,881 : -1</td>
<td>D</td>
<td>1488</td>
</tr>
<tr>
<td>RHTO_01538</td>
<td>SSE1</td>
<td>Chaperone</td>
<td>Adenyl-nucleotide exchange factor SSE1</td>
<td>Scaffold14: 28,485 - 29,984 : -1</td>
<td>T</td>
<td>506</td>
</tr>
<tr>
<td>RHTO_01660</td>
<td>ENO2</td>
<td>Sterol Metabolism</td>
<td>Farnesyl-pyrophosphate synthetase</td>
<td>Scaffold14: 346,860 - 348,381 : -1</td>
<td>T</td>
<td>297</td>
</tr>
<tr>
<td>RHTO_03043</td>
<td>FBA1</td>
<td>Glycolysis</td>
<td>Fructose-bisphosphate aldolase, Class II</td>
<td>Scaffold2: 906,477 - 908,296 : 1</td>
<td>T</td>
<td>373</td>
</tr>
<tr>
<td>RHTO_03322</td>
<td>LHS1</td>
<td>Chaperone</td>
<td>Hsp70 family chaperone LHS1</td>
<td>Scaffold20: 56,750 - 58,249 : 1</td>
<td>D</td>
<td>340</td>
</tr>
<tr>
<td>RHTO_03429</td>
<td>LHS1</td>
<td>Chaperone</td>
<td>Hsp70 family chaperone LHS1</td>
<td>Scaffold20: 325,239-326,738 : 1</td>
<td>T</td>
<td>339</td>
</tr>
<tr>
<td>RHTO_03430</td>
<td>NAF1</td>
<td>Pre-rRNA Processing</td>
<td>RNA-binding snrnp assembly protein</td>
<td>Scaffold20: 329137:330636:1</td>
<td>T</td>
<td>598</td>
</tr>
<tr>
<td>RHTO_03560</td>
<td>ACT1</td>
<td>Cytoskeletal</td>
<td>Actin</td>
<td>Scaffold21:244966:246465:1</td>
<td>D</td>
<td>782</td>
</tr>
<tr>
<td>RHTO_03746</td>
<td>TDI1</td>
<td>Glycolysis</td>
<td>Glyceraldehyde 3-phosphate dehydrogenase</td>
<td>Scaffold22: 315,363-317,069:1</td>
<td>T</td>
<td>1085</td>
</tr>
<tr>
<td>RHTO_03759</td>
<td>LHS1</td>
<td>Chaperone</td>
<td>Hsp70 Family Chaperone LHS1</td>
<td>Scaffold22:346190:347689:1</td>
<td>T</td>
<td>~1967</td>
</tr>
</tbody>
</table>

Abbreviations given in column one for each *R. toruloides* gene name as derived from *R. toruloides* NP11, accession GCA_000320785.2 (Zhu et al., 2012). *S. cerevisiae* S288c protein homologues in *R. toruloides* NP11 were identified using tBLASTn (Madden et al., 1996) and gene functions identified from UniProt (Bateman, 2019). Potential promoter transcriptional interference was determined by comparison between the *R. toruloides* gene of interest and the directionality of its immediate upstream gene (T, tandem; D, divergence).
Figure 16 | Strategy for generation of heat shock protein promoter (HSPP) fragment-containing reporter plasmid. A. Representative schematic diagram of HSPP primer locations relative to HSP in *R. toruloides* genome. Amplicons are produced using either a Forward (F) primer sequence specific to 1200 bp upstream of the HSP ORF (‘Fixed Length’ FL) or until the ORF of the immediately upstream gene (‘Intergenic’ IG). Both FL and IG primers are flanked with a 5’ 40 bp homology sequence to pEGFP-Rt-YR-Hyg base plasmid filler sequence, and a Reverse (R) primer containing a 5’ 40 bp homology sequence to eGFP. B. Representative schematic diagram of HSPP primer locations relative to HSP in base plasmid construct. Correct construction of plasmids p(01-FL)EGFP-Rt-YR-Hyg (top) or p(01-IG)EGFP-Rt-YR-Hyg (abbreviated to 01-FL and 01-IG respectively) was verified using specific flanking primers HSPP Check| F and HSPP Check| R.
4.2. Construction of HSPP reporter plasmids

Having generated insert amplicons containing HSPPs and homology regions to the base plasmid as described in section 4.1, pEGFP-Rt-YR-Hyg PGK promoter was excised through digestion with *AflII* and *PmlI* before yeast-based recombination in *S. cerevisiae* was performed using the method described in section 2.3.7 to make HSPP-driven EGFP reporter plasmids p(HSPP)EGFP-Rt-YR-Hyg.

Yeast colony PCR was employed to screen putative *S. cerevisiae* transformants using HSPP Check Primers as shown in Figure 16B. Where correctly-sized PCR products were amplified from colonies (both FL, containing an upstream gene, and IG containing just an intergenic region), plasmids were isolated and transformed into *E. coli* DH5α. Plasmids were isolated from *E. coli* transformants and the plasmids verified through restriction digest with *XbaI* and *ScaI*. Resultant linear fragments were size-fractionated in an agarose gel, as shown in representative example Figure 19. Verified plasmids were transformed into *Agrobacterium tumefaciens* as detailed in section 2.3.6 before transformation of *R. toruloides* with the plasmid t-DNA cassette through *A. tumefaciens*-mediated transformation, as detailed in section 2.4.
Figure 17 | PCR amplicons of heat shock protein promoters. Promoter fragments were amplified from extracted genomic DNA from *R. toruloides* CBS14 using primers as described in Table 4. (A) NEB Log2 Ladder was loaded into lanes labelled marker (M) with full-length amplicons loaded in odd-numbered lanes and intergenic-length amplicons loaded in even-numbered lanes: lanes 1-2 HSPP01; lanes 3-4 HSPP02; lanes 5-6 HSPP03; lanes 7-8 HSPP04; lanes 9-10 HSPP05; lanes 11-12 HSPP06; lanes 13-14 HSPP07; lanes 15-16 HSPP08; lanes 17-18 HSPP09; lanes 19-20 HSPP10; lanes 21-22 HSPP11; lanes 23-24 HSPP12; lanes 25-26 HSPP13.
Figure 18 | Example gel diagram of Yeast Colony-PCR verification of HSPPs. Three *S. cerevisiae* biological replicate colonies for each HSPP acted as template DNA for yeast colony-PCR, using flanking primers as described in Table 2, to verify desired HSPP-induced GFP plasmids for transformation into *R. toruloides* were constructed. Tracks 01 and 25 are loaded with MEB Log2 Ladder, with tracks 02-22 loaded with colonies theoretically containing 01-FL, 01-IG, 02-FL, 02-IG, 03-IG, 03-NC, 04-FL, where consecutive tracks contain one of each biological replicate respectively. Track 23 contains base plasmid as described in Figure 5 as a positive control, with track 24 as a negative control.
Figure 19| Representative gel of diagnostic plasmid digests of HSPP19-26. Plasmids extracted from *E. coli* containing putative HSPP-inducing GFP plasmids for transformation into *R. toruloides* containing *E. coli* before double digestion with both ThermoFisher FastDigest restriction endonucleases XbaI and Scal. NEB Log2 ladder was loaded into column M. Representative columns were loaded as follows: 9-FL (01-02), 11-FL (03-06), 11-IG (07-10), 12-FL (11-14), 13-FL (15-18), 13-IG (19-20), *PGK1*-1500 (23-24), with even numbered columns digested and odd numbered columns undigested plasmid negative controls. Anticipated lower band sizes of 1106, 2529, 2529, 2111, 2111, 2529, 2529, 2836, 2836 and >1967 (incomplete available sequence data) were expected for digested columns respectively; 9-FL band size is unexpected and a second transformation independent colony was then verified and used (data not shown).
4.3. Heat shock promoter induction in *R. toruloides*

To determine the potential use of *S. cerevisiae* HSPP homologues in *R. toruloides* as biotechnologically useful promoters (namely strongly inducible, and tightly regulated promoter activity), three biological *R. toruloides* replicates from each *A. tumefaciens* mediated HSPP-containing transformation were inoculated into synthetic medium (to limit the level of autofluorescence observed that would otherwise be present in other media such as YPD) and heat-shocked as described in section 2.9. Briefly, late logarithmic-phase yeast cultures were diluted to an OD$_{600}$ 0.2, given an hour to recover at 25 °C followed by either a heat shock to 37 °C or remaining at the control temperature of 25 °C, before returning to the initial temperature for five hours’ growth, and the fluorescence of the yeast samples measured through flow cell cytometry.

Induction results are shown in Figure 20, where fluorescence of >16 000 cells in each group is shown for both control (blue) and heat shocked (red) conditions. Of the HSPP-driven GFP-containing yeasts, an apparent difference from wildtype (WT) is found in HSPP 1-FL (*PGK1*), 2-FL (*ENO2*), 4-IG (*TPI1*), 5-FL (*SSE1*), 6-FL (*ERG20 – heat shocked only*), 11-FL (*ACT1*), and 12-FL (*TDH3*). From these transformants no statistical difference between heat shock and control conditions were identified, but 2-FL appears to show a possible difference between control and heat shock conditions which may not have been identified statistically due to the large variance in fluorescence and deserves further analysis. This suggests the identification of a further six promoters which appear constitutively expressed in *R. toruloides* and require further characterisation. It is curious that HSPP1, which is also *PGK1* promoter albeit at a cut-down of 1200 bp, had an apparent difference in fluorescence against the *PGK1*-1500 Johns *et al.*, (2016) promoter and suggests further analysis is needed here.
**Figure 20** | Induction of heat shock protein promoters controlled cell eGFP fluorescence. Median pixel intensities of logarithmic-stage *R. toruloides* cultures after dilution to an OD$_{600}$ of 0.2 before either shock-treatment to 37 °C or remaining at 25 °C for an hour followed by a five-hour recovery period at 25 °C before: paraformaldehyde-fixation; excitation at 488 nm; and photography with an Amnis ImageStream X Mark II Imaging Flow Cytometer. X-axis numbers represent the HSPPs denoted in Table 5, where >16 000 images from three biological replicates of each HSPP-EGFP of each were treated; error bars represent Median Absolute Deviation. Blue and red crosses denote an identified statistical difference between HSPP and WT fluorescence medians respectively using a Student’s *t*-test *p*=0.05. No statistical differences were identified between HSPPs under heat-shock conditions.
4.4. Heat Shock promoter kinetics in *R. toruloides*

Having initially measured the final HSPP-driven fluorescence at six hours, it was decided to measure promoter kinetics from the initial heat-shock in case an induction peak was being temporally missed. In *S. cerevisiae*, RNA transcription is upregulated for all of the selected HSPs within 5 minutes, with the majority of HSPs' RNA transcripts reaching maximal expression between 15-20 minutes. As detailed in section 2.10, logarithmic-stage *R. toruloides* was normalised to an OD$_{600}$ of 0.2 in a 24-well plate; both OD and fluorescence were then measured at regular intervals during a two hour heat shock to 37 °C, and the following four hour recovery time at 25 °C (Causton *et al.*, 2001; Gasch, A. P. *et al.*, 2000). Although unlikely, GFP stability in *R. toruloides* has not yet been determined experimentally and it is probable that the five promoters identified in section 4.3 are constitutively expressing GFP, thereby replenishing degraded fluorescent protein, whereas in a briefly-expressed RNA transcript, such as that only produced while Msn2/4 is nucleus-localised while hyperphosphorylated, the unstable fluorophore would not be detectable after degradation and non-replenishment.

Kinetic assays of the promoters HSPP 1-FL (*PGK1*), 2-FL (*ENO2*), 2-IG (*ENO2*), and 12-FL (*TDH3*) are shown in Figure 21. Statistical differences were determined by Student’s *t*-test, and differences between WT and HSPP-induced GFP under heat-shock are identified by red crosses; differences between WT and HSPP-induced GFP when remaining at 25 °C for the duration of the experiment are identified in blue crosses; and differences in fluorescence between HSPP-GFP at heat shock and control 25 °C are represented by black crosses. Curiously, 02-FL showed a rapid accumulation of fluorescence during recovery, reaching maximum fluorescence after 50 minutes. This is statistically different to HSPP-driven GFP under control conditions, suggesting the presence of an inducible heat-shock response. However, the HSPP appears to “leak”, as a weaker control fluorescence constantly increased, becoming not statistically distinguishable from the fluorescence level of the heat-shocked transformants after 4.5 hours. Interestingly, heat-shocked WT also shows a ‘bounce’ of fluorescence shortly after heat shock; this is potentially due to autofluorescence as part of the ESR. Even more curiously, 02-IG HSP-driven GFP also continuously increases, as for 02-FL. However, this transformant does not exhibit an inducible heat-shock fluorescence-associated response, therefore suggesting the loss of a crucial
stress-response element contained within the upstream coding DNA sequence. Interestingly, 12-FL does not show significant difference between WT and HSPP-transformants until after 5 hours; this suggests the presence of a weak constitutive promoter. This data corroborates the fluorescence characteristics determined for this promoter in Figure 20, where stronger fluorescence was observed than for WT 6 hours after heat shock.

These kinetic experiments suggest that it may be prudent for future experiments on *R. toruloides* seeking to identify inducible HSPPs to focus primarily on the first 2-4 hours after heat shock, after which any sudden inducibility may be masked. To this end, kinetic assays of fluorescence of all HSPPs (FL and IG) were undertaken, but no statistical differences were observed between the other HSPPs and WT in these cases (data not shown).

This experiment does not appear to corroborate findings by Johns *et al.*, (2016) and Wang, Y. *et al.*, (2016), with regards to the strongly constitutive promoter *PGK1*. Through this time-course experiment, no statistical difference between WT and *PGK1*-controlled GFP could be observed, despite it being a well-characterised and strong constitutive promoter. However, a trend can be observed where, although WT continues to stably not fluoresce, a continuous upward trend in fluorescence can be seen for both the heat-shocked and non-heat-shocked HSP-transformant. This may indicate that other ‘HSP’-transformants may not be biologically useful heat-shock inducible protein promoters but may be constitutive promoters which have not yet been identified due to truncation of the experimental kinetic time series at 6 hours. It may, therefore, be prudent to extend this time course in the future to allow possible constitutive promoters, like *PGK1*, to be identified.

Possible reasons for the low induction in all chosen *S. cerevisiae* HSPP homologues in *R. toruloides* are numerous and, from this study, no definitive conclusions can be drawn. However, one likely cause is that the native ESR was not adequately induced in this organism during heat shock to 37 °C because CBS14 strain of the yeast may be ordinarily tolerant to short stresses found at 37 °C as reported in CBS349 by Zhu *et al.*, (2012b). To determine CBS14 strain thermotolerance in the future WT yeast cultures could be temporarily heat-shocked at increasing temperatures before plating until a maximum viable temperature is found. As outlined in section 1, the oleaginous yeast natively produces compounds which are capable of actively mitigating damage
from heat shock-derived deleterious compounds (developing 940.2 μg/g carotenoids in *R. toruloides* strain NP11 (Lin *et al.*, 2017)) and capable of sequestering reactive oxygen species generated during heat shock. Additionally, the physiological presence of a large lipid droplet in *R. toruloides*, which sequesters misfolded proteins in the ascomycete yeast *S. cerevisiae*, may have prevented the necessity for additional molecular chaperones and the associated additional transcriptional- and translational-energy burden of up-regulating these proteins under a ‘milder’ heat shock (Henne *et al.*, 2019).

It is possible that a stronger induction of HSPP-driven heterologous genes could be found in thermotolerant strains of the yeast. Wu *et al.*, (2018) employed an adaptive breeding strategy to selectively drive and isolate a thermotolerant *R. toruloides* TK16 strain which maintained viability when grown at 37 °C, whereas WT growth was attenuated above 36 °C. Indeed, dry biomass was comparable with higher titres of lipid produced at 37 °C between the mutant strain when compared to WT at 30 °C. If the mechanism of acquired *R. toruloides* thermotolerance mirrors that of *S. cerevisiae*, where heat shock proteins are more strongly up-regulated (Satomura *et al.*, 2013, 2016), then it is possible that a stronger and more sustained HSPP-driven response could be identified.
Figure 21: Graphical series of measured fluorescence of heat shock protein promoters (HSPP) controlled eGFP after heat shock. Time-dependent response of EGFP fluorescence of a representative tested HSPPs (as annotated in Table 5) after excitation at 488 nm. As shown in the top graph (red), after 5 min the temperature of the ClarioStar platereader rises to 37 °C before cooling back to 25 °C after 2 hours; the negative control (blue) remains constantly at 25 °C. The remaining 4 graphs show OD₆₀₀ normalised fluorescence of the HSPP-GFP (red and blue) and WT *R. toruloides* yeast (orange and green) after either heat-shock or constant temperature respectively. Black crosses represent a statistical different between HSPP-EGFP containing yeast under heat-shock or negative control conditions; red crosses represent a statistical difference observed between HSPP-EGFP containing and WT yeast under heat-shock conditions; and blue crosses represent a statistical difference observed between WT and HSPP-EGFP containing yeast under negative control conditions.
5. Conclusion

As the crisis of climate change becomes more apparent, it is self-evident that a change in global methods for reducing anthropogenic carbon emissions is urgently required. While inroads in other renewable energy technologies capable of eventually fulfilling static and small-distance energy requirements through low-density batteries can be fulfilled, there is not yet a current feasible alternative to the mobile energy density fuels necessitated by longer distance transport than petrochemicals. Recent progress in developing *R. toruloides* as a biotechnological chassis for high-titre neutral lipid synthesis and accumulation from low-cost organic matter, which can undergo transesterification to drop-in FAME biodiesel, offers an alternative to continued release of sequestered carbon as greenhouse gases. Additionally, *R. toruloides* has demonstrated increased or exotic production of high-value lipids and products through expression of foreign genes, including carotenoids, used in food colourants, and essential lipids for dietary supplements or as a therapy for certain neurological and metabolic pathologies.

We have demonstrated the first use of fluorescently epitope-tagged protein probes for live-cell imaging lipid droplets in *R. toruloides*. This can now be shown to quantify lipid biogenesis in this oleaginous yeast and can be used to monitor and investigate mechanisms of lipid droplet formation and maintenance in the yeast in conjunction with the ER, which we have also putatively tagged with the GFP fluorophore. To aid characterisations between the dynamics between the two, different fluorophore tags, such as RFP or Venus yellow, will have to be synthesised. Furthermore, the autophagy complex has been targeted, but the specificity of this tag requires further verification.

Having set out to identify heat-shock inducible promoters for genetic manipulation in undefined media, we instead firmly identified five constitutive promoters. However, one of these identified promoters (RHTO_00323; ENO2) is also potentially heat-shock inducible. However, at their current level of characterisation, each of these promoters requires further experimentation to become a biotechnologically useful genetic tool. We have, however, indicated that the rate of heat-shock in *R. toruloides* is rapid, and further experiments to characterise and identify HSPPs should focus on the immediate two hours after heat shock to identify differences in protein expression.
6. Acknowledgements

In the first instance, I am eternally grateful to my academic supervisor Dr Stephen Aves for his long-standing support, patience, and assistance in all aspects of this project, from the initial experimental designs to troubleshooting to suggesting changes to this manuscript.

I am also grateful for the companionship, technical expertise and moral support from my colleagues in the Mezzanine Laboratory and beyond in the Biosciences department, namely Francesco Valente, Stefan Sassmann, Alex Johns, Tom Chaloner, William Kay, Corey Holt, Dominic Wiredu Boakye, Andy Foster, Anke Lange, Jamie Gilman, Skye Marshall, Nick Gaunt, Peggy Dousseaud, Richard Tennant, Paul James, Ann Power, Katy Jones, Maulood Turfah, Dominic Arnaud, Sheera Abdulla and Chloe Singleton.

Finally, I should like to thank my parents, Richard and Sue, for their continued encouragement and belief in me.
7. Bibliography


Ballio, A., Di Vittorio, V., and Russi, S. (1964) The isolation of trehalose and polyols


Bose, S., Dutko, J. A., and Zitomer, R. S. (2005) Genetic factors that regulate the


Herms, A., Bosch, M., Ariotti, N., Reddy, B. J. N., Fajardo, A., Fernández-Vidal, A.,


Klionsky, D. J. (2011) For the last time, it is GFP-Atg8, not Atg8-GFP (and the same goes for LC3). Autophagy 7(10): 1093–1094.


Macool, D. J., Zhixiong, X., and Quinn Qun, Z. (2013/March) Expression of cytosolic malic enzyme in transgenic yarrowia to increase lipid production.


Naik, S. N., Goud, V. V., Rout, P. K., and Dalai, A. K. (2010) Production of first and
second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews*.


Sztalryd, C., and Brasaemle, D. L. (2017) The perilipin family of lipid droplet proteins: Gatekeepers of intracellular lipolysis. *Biochimica et Biophysica Acta - Molecular and


