The function of the *TLO* gene family of *Candida albicans*

Jessica Fletcher

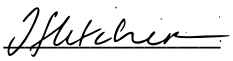
April 2022

A thesis submitted to the University of Dublin, Trinity College in fulfilment of the requirements for the degree of Doctor of Philosophy

Microbiology Research Unit
Division of Oral Biosciences
Dublin Dental University Hospital
Trinity College Dublin

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement. I consent to the examiner retaining a copy of the thesis beyond the examining period, should they so wish (EU GDPR May 2018).



Jessica Fletcher

Summary

Candida albicans is a commensal yeast species and a member of the normal flora of the gastrointestinal and vaginal tract in humans. It is also associated with opportunistic infections, primarily in patients who are immunocompromised. The pathogenesis these infections is complex, however, it has been proposed that the expansion of specific virulence factor gene families in *C. albicans* contributes to its increased virulence. One such family is the *TLO* gene family. *Candida albicans* strains can possess 10-15 of these telomere associated ORFs, *C. dubliniensis*, two, and all other *Candida* species have a single homologue. The *C. albicans TLOs* are split into three clades, α, β and γ, based on their C-terminal domains and data suggests they may have developed individual functions through their expansion. The Tlo proteins are Med2 homologues, incorporating into the tail of the Mediator complex, and are thought to play an important role in controlling transcription. It has also been proposed that there is a free pool of Tlo protein in *C. albicans,* not associated with Mediator (not present in *C. dubliniensis*), which is believed to have a function independent of the complex. The aim of this thesis was to investigate the role of the *TLO* gene family in *C. albicans*.

Experiments were designed to express clade representative *C. albicans TLO*s in the *C. dubliniensis* WT under either a native or a strong promoter to investigate if expansion of the *TLO* repertoire in this species could enhance virulence. Phenotypic assays revealed that these *C. dubliniensis* strains were more tolerant of stress, more able to form biofilm and were more virulent than the parental strain, confirming the hypothesis that the expansion of the *TLO* family is likely important for increased virulence.

A *TLO* knockout strain was generated in *C. albicans* using CRISPR-Cas9 mutagenesis, targeting all 14 diploid family members in one transformation. Routine PCR and WGS confirmed deletion of all *TLO*s. Phenotypic testing highlighted significant defects in cellular morphology (pseudohyphal compared to WT blastospores), growth rate, stress tolerance, biofilm formation and virulence in the *ΔΔtlo* mutant, indicating that *TLO* genes are essential for fitness and virulence.

Single, clade representative *TLO* genes were then reintroduced into the *ΔΔtlo* background to investigate the roles of each gene. Introduction was confirmed by PCR, and qRT PCR confirmed mRNA expression from all constructs. Western blot analysis confirmed expression of protein from the constructs, apart from those expressing

TLOγ11, possibly due to sequence specific targeted degradation of this protein. Phenotypic testing highlighted that introduction of *TLOα1* into the *ΔΔtlo* background could restore cell morphology, growth rate, stress tolerance, biofilm formation and virulence to WT levels. While reintroduction of *TLOβ2* restored growth rate, biofilm formation and virulence, as well as some level of tolerance to stress. The reintroduction of *TLOβ2* also restored the WT morphology, while also generating a subpopulation of hyphal cells in non-hypha-inducing conditions. The reintroduction of *TLOγ11* did not appear to have any effect on any of the phenotypes tested.

RNA-seq analysis was performed to examine the effects of deletion of the *TLO* family and of the expression of individual *TLO* genes in the *ΔΔtlo* background. ChIP-seq analysis was performed to determine where the Tlo proteins were interacting with DNA within the cells. Data from these experiments suggested that the *TLO* genes play key roles in many important pathways in the cell. The *ΔΔtlo* mutant was deficient in expression of genes related to carbohydrate metabolism, as well as genes important for cellular morphology and cell membrane composition. The *ΔΔtlo* was also found to express genes typically upregulated upon exposure to external stress and antifungal drugs. Reintroduction of *TLOα1* and *TLOβ2* restored transcription of many of the genes related to carbohydrate metabolism and morphology, and resolved aberrant stress responses. It appeared that the interaction of Tlos with genes encoding transcription factors, e.g. *TYE7*, and with other genes in these pathways was responsible for the restoration of mutant phenotypes. While the Tloα1 and Tloβ2 proteins were found to localise to many of the same sites, unique interactions were also discovered, such as Tloα1 interacting with biofilm formation genes, and the Tloβ2 protein interacting hyphal morphogenesis genes. Reintroduction of *TLOγ11* did not have a significant effect on the transcriptome, however the Tloγ11 protein was found to localise to telomeres and repeat sequences, which may indicate a role for this protein in maintaining chromosome stability.

This project has shown that expansion of the *TLO* gene family can increase fitness and virulence of *Candida* spp., that deletion of the *TLO* gene family in *C. albicans* results in a less fit and less virulent strain, and that the reintroduction of specific *TLO*s can restore virulence attributes. Investigation into the role of each *TLO* through NGS found that individual Tlo proteins may have distinct roles within transcriptomic pathways, and can regulate fungal virulence. In all, this work has highlighted that the expansion of the *TLO* gene family in *C. albicans* is an important feature may have contributed to its enhanced capacity to colonise and cause disease in humans.

Acknowledgements

Firstly, I would like to thank my supervisors, Professor Derek Sullivan and Professor Gary Moran, for their support and guidance throughout this process. Without the opportunities and resources they provided, this work would not have been possible. I would also like to thank Professor David Coleman, who offered his expertise when needed, and Dr Mary O'Donnell for being an excellent lab manager. I am very grateful to have carried out my research in such a stimulating and encouraging environment.

I would also like to acknowledge Dublin Dental University Hospital and Trinity College Dublin for providing the financial support that funded this research.

To my friends in the lab, I would like to thank you all for being such great co-workers over the past four years, for all the tea-times and for all the after work socials. To Dr Emily Deasy, Dr Peter Kinnevey and Dr Brenda McManus for sharing their experience and for being on hand to help answer any question, no matter how trivial. To Dr Peter Flanagan for taking me under his wing, and for not playing the Murder, She Wrote theme song on my first day in the lab. To Dr James O'Conner-Moneley for all his postdoc support, and for always being there when I needed to vent. To Dr Ajith Kumar Selveraj and Claire Crowley, and to Dr Dean Frawley and Melvin Wen Jun Lim for being such great team mates, and for sharing their research with me at our group meetings. To Bisola Aloba, Liam Grealy and Nicole Kavanagh for their camaraderie. To Michaela Kearney for eventually becoming friends with us down in the basement. To Dr Megan Earls for being a great friend, and an international woman of mystery. I am so happy to have met you all.

A very special thank you to Dr Sarah Egan and Dr Elaine Moloney, for being such amazing people. For the times we laughed until we cried, and cried until we laughed. I am very lucky to have you both as friends.

Thank you to all my friends who were cheering me on along the way, who were always there for moral support or a night out. To Siobhán Barber, Amanda Nelson and Rachel Nelson, now might be time for that reunion tour we've always talked about. To Cara Donaghey and Clíodhna Gallagher, two of my oldest friends. To Cathal McKinney, Katy Finnegan and Aishling Mohan for the dinners and dancing. And to the members of Na Gaeil Aeracha, for giving me an outlet for all the thesis stress both on and off the pitch.

To Louise Keogh, for being there at the end of every long hard day, thank you for everything. Thank you also to Louise's parents Dave and Deirdre.

Finally, to my own family. To my parents, John and Janet and to Shaun and Rachel, thank you for supporting me through everything. For helping me to pursue my goals and giving me every opportunity, thank you. Thank you to my Granny Eileen Fletcher and my great aunt Rosaleen Hegarty for being such inspirational women to look up to. And yes Granny, I'm finally finished college.

Table of Contents

List of Figures

List of Tables

Abbreviations

AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
AMP	Adenosine monophosphate
approx.	Approximately
Appx.	Appendix
ATP	Adenosine triphosphate
BMDM	Bone marrow derived macrophages
bp	Base pair(s)
BTS	Bermuda Triangle sequence
BWA-MEM	Burrows Wheeler Aligner - Maximal Exact Match
cAMP	Cyclic adenosine monophosphate
cDNA	complementary DNA
cfu	Colony forming unit
CFW	Calcofluor White
CGD	*Candida* Genome Database
CHEF	Clamped homogenous electric fields
ChIP	Chromatin immunoprecipitation
ChIP-seq	ChIP followed by sequencing
cm	Centimetre
CR	Congo Red
CRISPR	Clustered regularly interspaced short palindromic repeats
crRNA	CRISPR derived RNA
C_T	Threshold cycle
CTD	C-terminal domain
DDUH	Dublin Dental University Hospital
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxyribonucleotide triphosphate
DOX	Doxycycline
dsDNA	Double stranded DNA
dT	Deoxythymine
DTT	Dithiothreitol
e.g.	*Exempli gratia*; for example
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid
et al.	*Et alia*; and others
etc.	*Et cetera*; and the rest
FBS/FCS	Foetal bovine serum/Foetal calf serum
g	Gram
GO	Gene ontology
gRNA	Guide RNA
GSEA	Gene set enrichment analysis

h	Hour(s)
H_2O	Water
H_2O_2	Hydrogen peroxide
HA	Hemagglutinin
HAGs	Hyphal associated genes
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV	Human immunodeficiency virus
HS	Heat shock
i.e.	*Id est*; that is
IC	Invasive candidiasis
ICU	Intensive care unit
IP	Immunoprecipitate
IVF	*In vitro* fertilisation
kV	Kilovolts
L	Litre
L broth	Lysogeny broth
L_2FC	Log_2 fold change
LTR	Long terminal repeat
M	Molar
mA	Milliampere
MAP	Mitogen-activated protein
MAPK	MAP kinase
mg	Milligram
MIC	Minimum inhibitory concentration
min	Minute(s)
ml	Millilitre
mM	Millimolar
MNase	Micrococcal nuclease
mRNA	Messenger RNA
MRS	Major repeat sequence
MW	Molecular weight
NAD	Nicotinamide adenine dinucleotide
NAT	Nourseothricin
NAT200	200 μg/ml nourseothricin
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
ng	Nanogram(s)
NGS	Next generation sequencing
nm	Nanometre
OC	Oral candidiasis
OD	Optical density
ONT	Oxford Nanopore Technology
OPC	Oropharyngeal candidiasis

ORF	Open reading frame
OS	Osmotic stress
PAM	Protospacer adjacent motif
PBS	Phosphate-buffered saline
PC	Personal computer
PCIA	Phenol:chloroform:isoamyl alcohol
PCR	Polymerase chain reaction
PDVF	Polyvinylidene difluoride
pH	Potential hydrogen
PI	Protease inhibitor
PIPES	Piperazine-N,N′-bis(2-ethanesulfonic acid)
post hoc	After this
qRT-PCR	Quantitative real time PCR
RHE	Reconstituted human epithelium
RNA	Ribonucleic acid
RNA-seq	Sequencing of RNA
RNAP	RNA Polymerase II
RNase	Ribonuclease
ROS	Reactive oxygen species
rpm	Revolutions per minute
RVVC	Recurring vulvovaginal candidiasis
SDS	Sodium dodecyl sulphate
SDS-PAGE	SDS–polyacrylamide gel electrophoresis
SHERLOCK	Specific High-Sensitivity Enzymatic Reporter UnLOCKing
SOC	Super optimal broth with catabolite repression
spp.	Species
TAD	Transcriptional activation domain
TAGEN	Telomere-adjacent gene expression noise
TBE	Tris-borate EDTA
tBOOH	*tert*-Butyl hydroperoxide
TCD	Trinity College Dublin
TE	Tris-EDTA
TPE	Telomere positional effect
TRE	*TLO* recombination event
tRNA	Transfer RNA
U	Units
UAS	Upstream regulatory sequences
USD	United States Dollar
UV	Ultra violet
V	Volts
v/v	Volume per volume
VVC	Vulvovaginal candidiasis
w/v	Weight per volume

WGD	Whole genome duplication
WGS	Whole genome sequencing
WT	Wild-type
x g	Times gravity
XS	Xenobiotic stress
YEP-Gal	Yeast extract peptone galactose
YEPD	Yeast extract peptone dextrose
YNB	Yeast nitrogen base
µg	Microgram(s)
µl	Microlitre(s)
µM	Micromolar
µm	Micrometre
%	Percentage
′	Prime
°C	Degree(s) Celsius
<	Less than
>	Greater than

Publication

PLOS ONE

RESEARCH ARTICLE

Expansion of the *TLO* gene family enhances the virulence of *Candida* species

Peter R. Flanagan[1,2☯], Jessica Fletcher[1,2☯], Hannah Boyle[1,2], Razvan Sulea[1,2], Gary P. Moran[1,2‡*], Derek J. Sullivan[1‡*]

1 Microbiology Research Unit, Division of Oral Biosciences, Dublin Dental University Hospital, Dublin, Ireland, **2** University of Dublin, Trinity College Dublin, Dublin, Ireland

☯ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* Derek.Sullivan@dental.tcd.ie (DJS); Gary.Moran@dental.tcd.ie (GPM)

OPEN ACCESS

Citation: Flanagan PR, Fletcher J, Boyle H, Sulea R, Moran GP, Sullivan DJ (2018) Expansion of the *TLO* gene family enhances the virulence of *Candida* species. PLoS ONE 13(7): e0200852. https://doi.org/10.1371/journal.pone.0200852

Editor: Janet Quinn, Newcastle University, UNITED KINGDOM

Received: May 14, 2018

Accepted: June 11, 2018

Published: July 20, 2018

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The project was supported by the Dublin Dental University Hospital and Science Foundation Ireland (11/RFP.1/GEN/3042). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abstract

The *TLO* genes are a family of subtelomeric ORFs in the fungal pathogens *Candida albicans* and *C. dubliniensis* encoding a subunit of the Mediator complex homologous to Med2. The more virulent pathogen *C. albicans* has 15 copies of the gene whereas the less pathogenic species *C. dubliniensis* has only two. To investigate if expansion of the *TLO* repertoire in *C. dubliniensis* has an effect on phenotype and virulence we expressed three representative *C. albicans TLO* genes (TLOβ2, *TLOγ11* and TLOα12) in a wild type *C. dubliniensis* background, under the control of either their native or the *ACT1* promoter. Expression of *TLOβ2* resulted in a hyperfilamentous phenotype, while overexpression of *TLOγ11* and *TLOα12* resulted in enhanced resistance to oxidative stress. Expression of all three *TLO* genes from the *ACT1* promoter resulted in increased virulence in the *Galleria* infection model. In order to further investigate if individual *TLO* genes exhibit differences in function we expressed six representative *C. albicans TLO* genes in a *C. dubliniensis Δtlo1/Δtlo2* double mutant. Differences were observed in the ability of the expressed *CaTLO*s to complement the various phenotypes of the mutant. All *TLO* genes with the exception of *TLOγ7* could restore filamentation, however only *TLOα9*, γ*11* and α*12* could restore chlamydospore formation. Differences in the ability of *CaTLO* genes to restore growth in the presence of H_2O_2, calcofluor white, Congo red and at 42˚C were observed. Only *TLOα3* restored wild-type levels of virulence in the *Galleria* infection model. These data show that expansion of the *TLO* gene family in *C. dubliniensis* results in gain of function and that there is functional diversity amongst members of the gene family. We propose that this expansion of the *TLO* family contributes to the success of *C. albicans* as a commensal and opportunistic pathogen.

Introduction

Candida species (spp.) are an important component of the human microbiota. They are found in a wide range of anatomic niches, particularly in the gastrointestinal and vaginal tracts and if host conditions provide an opportunity they can evade immune responses and cause a

spectrum of diseases, ranging from superficial infections of the mucosae to life threatening systemic infections in severely immunocompromised patients. In particular, *Candida* spp. have been cited as the fourth most common cause of nosocomial bloodstream infections [1].

The most pathogenic *Candida* species are *C. albicans*, *C. parapsilosis*, *C. tropicalis* and *C. glabrata*. [2]. *C. albicans* is by far the most commonly identified cause of candidiasis and is often regarded as the most pathogenic fungal species in humans. *C. albicans* is a highly versatile microorganism that has the ability to activate rapid transcriptional responses in order to adapt to changing environmental conditions, potentially allowing it to colonise and infect multiple anatomic sites [3].

One of the reasons why *C. albicans* is more pathogenic than other *Candida* spp. is that, apart from *Candida dubliniensis*, it is the only truly dimorphic *Candida* species, having the ability to switch between yeast and filamentous forms of growth [4]. Hyphal cells and hypha-specific proteins are well documented contributors to virulence and facilitate adherence to the host, penetration of tissues and the formation of biofilms [5–7].

Candida dubliniensis, which was first identified in 1995 [8] is very closely related to *C. albicans*. The genomes of both species are highly similar (i.e. 98% synteny [9]) and they share many phenotypic traits, including the capacity to form hyphae. Despite their very close relationship *C. dubliniensis* has been shown to be far less virulent than *C. albicans* and is only rarely found to be the cause of systemic infections [10]. Comparison of the genomes to identify the underlying genetic differences for the disparity in virulence found gene family size as the primary differentiating feature. Some of these differences were found in gene families known to contribute to virulence (e.g. the agglutinin-like sequence (*ALS*) and secretory aspartyl proteinase (*SAP*) gene families), however, one of the greatest differences was observed in the composition of the *TLO* (TeLOmere-associated) family in each species. The *C. albicans* SC5314 genome contains 15 *TLO* genes compared with just two in *C. dubliniensis* [9]. In *C. albicans*, as the name suggests, the *TLO* genes are situated close to the telomeres of each chromosome. The *C. albicans TLO* genes can be divided into four distinct clades based on the structure of their genes (Fig 1A). These include the relatively highly expressed α clade containing six members, a single β clade gene, the γ clade containing seven members, and the ψ clade containing a single pseudogene member [11–13]. *TLO* copy number varies between strains (10–15), with variation in the number of α and γ clade genes [14]. However, a single β clade gene is present in all genomes analysed to date [14]. The N-terminus of the *C. albicans TLO* genes encodes a conserved Med2 domain and these genes are now known to encode the Med2 component of the Mediator complex [11,15]. Mediator is a large multi-subunit protein complex which is conserved throughout eukaryotes and mediates interaction between RNA polymerase II and the machinery used in the initiation of transcription at target gene promoters [16,17]. Uwamahoro *et al.* (2012) demonstrated that *C. albicans* Mediator has a role in the expression of genes related to virulence traits [17] and it has also recently been demonstrated to play a role in resistance to antifungal drugs [18,19]. Tlo/Med2 forms a part of the Tail module of Mediator, along with Med3 and Med15 [15]. Different Tlos are found at different levels in *C. albicans* and *C. dubliniensis* [15] suggesting that there are pools of Mediator in each species with a different Med2 component. Given the size of the *TLO* family in *C. albicans* it has also been proposed that there is a substantial pool excess of "Mediator-free" Tlo in this species [11].

Deletion of the two *TLO* genes in *C. dubliniensis* resulted in defects in activation of transcriptional responses associated with a number of virulence traits including tolerance of oxidative stress and hypha formation [12], while overexpression of *CdTLO2* (and creation of a pool of "free" Tlo), but not *CdTLO1*, in *C. dubliniensis* results in hyperfilamentation [20]. As well as confirming the role of Med2 in virulence, the data from *C. dubliniensis* suggest differences in functionality amongst the two Tlo proteins expressed in that species.

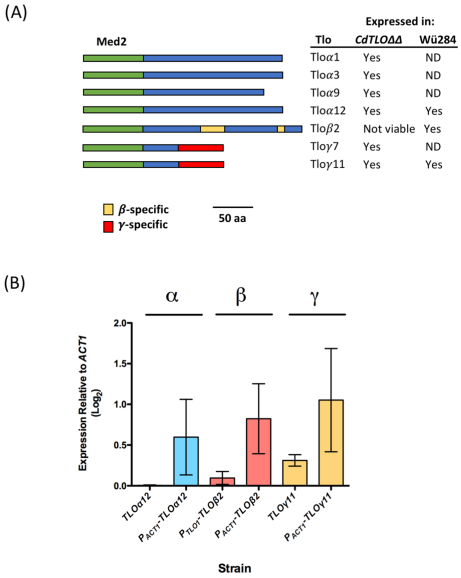


Fig 1. Structure and expression of *C. albicans TLO* genes. (A) Diagram comparing the structure of the Tlo proteins analysed in this study, based on the models of Anderson *et al.* [11]. The green box represents the conserved Med2-like domain. The blue box represents the clade-specific c-terminus. The γ- and β-specific regions are indicated by yellow and red boxes, respectively. The table on the right indicates which genes have been expressed in wild-type *C. dubliniensis* Wü284 and the *TLO* null derivative *ΔΔtlo*. ND = not done. (B) RT-PCR expression data of *C. albicans TLO*s expressed in the *C. dubliniensis* WT Wü284 strain. RT-PCR expression graphs represent the results of three independent experiments.

https://doi.org/10.1371/journal.pone.0200852.g001

We have previously proposed that the increased virulence of *C. albicans* compared to other *Candida* species may be due to an increased transcriptional flexibility due to its expanded family of Tlo proteins which may have differences in functionality [12]. Evidence that individual Tlo proteins have specific function(s) in *C. albicans* has recently been provided by Dunn *et al.* [21] who investigated the phenotypic effect of controlling the expression of individual *TLO* genes using a Tet-ON system. In order to investigate our hypothesis, we have added to the repertoire of *TLO* genes in *C. dubliniensis* by heterologously expressing representative *C. albicans TLO* genes in a wild-type *C. dubliniensis* strain. The rationale for selecting *C. dubliniensis* as the host species was due to the low copy number of native *TLO* genes in this species as we reasoned that it could be difficult to determine a phenotypic effect in a *C. albicans* background of 15 *TLO* genes. In addition, we also attempted to investigate the functional diversity within the *C. albicans TLO* gene family, by expressing *TLO* representative *CaTLO* genes in a *C. dubliniensis ΔΔtlo* double mutant and identified the phenotypes conferred by each gene. We propose that our data demonstrate possible evolutionary advantages associated with *TLO* gene duplication and diversification.

Materials & methods

Candida strains & culture conditions

The strains of *Candida* spp. used in this study and their genotypes are listed in S1 Table. All *Candida* strains were routinely grown on Yeast Extract Peptone Dextrose (YEPD) agar at 37˚C. Nourseothricin-resistant transformants were cultured on YEPD agar containing nourseothricin [100 μg/ml (NAT100)]. Lee's Medium [22] and Spider medium [23] were used to induce filamentation. Cornmeal agar supplemented with 1% Tween® 80 was used for chlamydospore formation.

For spot plate assay experiments, a suspension of 2 x 10^6 cells/ml was prepared from overnight cultures and 7 μl from serial dilutions (10^0 to 10^{-4}) were spotted onto YEPD agar plates containing the indicated agents. The plates were incubated at 37˚C for 48 h in a static incubator. Growth was recorded using a Flash n' Go plate visualizer (IUL Instruments). Each experiment was carried out on three separate occasions.

For liquid culture, YEPD broth was used in an orbital incubator at 200 r.p.m. at the indicated temperature. In order to determine the doubling times of strains, the optical densities of cultures were measured at 600nm during the exponential growth phase and plotted using Prism GraphPad (GraphPad, CA, USA). Doubling times were calculated from 3 replicate growth curves. Galactose (2% w/v) was substituted for glucose where indicated. Induction of filamentation in liquid cultures was carried out with cells from overnight YEPD broth cultures grown at 30˚C, which were washed twice with sterile Milli-Q water (Millipore Ireland B.V., Co. Cork, Ireland) and added to hyphal-induction medium (10% v/v foetal bovine serum in dH_2O) to a density of 2 x 10^5 cells/ml in a six-well tissue culture plate at 37˚C. The numbers of true hyphal cells were quantified using a Nikon E600 microscope and a Nikon TMS-F inverted light microscope (Nikon U.K., Surrey, U.K.). Experiments were carried out on three separate occasions.

Heterologous expression of *C. albicans TLO* genes in *C. dubliniensis*

C. albicans TLO genes β*2*, γ*11* and α*12* were heterologously expressed in the *C. dubliniensis* wild type strain Wü284. *TLO*γ*11* and α*12* were expressed under the control of their native promoters, however the *TLO*β*2* promoter sequence is incomplete in the SC5314 genome sequence so a fusion gene with the *TLO*α*1* promoter (S1 File) was synthesised by GeneWiz (Essex, UK) and inserted in the *Xho*I and *Hin*dIII restriction endonuclease sites of pCDRI [24]. *TLO*γ*11* and α*12* were amplified from SC5314 using gene-specific primers (S2 Table) containing recognition sequences for *Xho*I and *Hin*dIII restriction endonucleases. Digested amplimers were ligated to *Xho*I/*Hin*dIII cut pCDRI plasmid using T4 DNA ligase (Promega, Wisconsin, USA) and transformed into *E. coli* XL10 competent cells (Sigma-Aldrich, Missouri, USA) using standard protocols. Transformants were selected on pre-warmed Lysogeny (L) agar supplemented with 100 μg/ml ampicillin. Plasmid pCDRI and its derivatives were linearised with *Eco*47III and transformed in *C. dubliniensis* as described by Staib *et al.* [25]. Transformants were selected on YEPD agar containing 100 μg/ml nourseothricin. Additional plasmid constructs containing *TLO*s β*2*, γ*11* and α*12* were also generated in pGM161, which is a derivative of pCDRI allowing expression from the *ACT1* promoter, using the same cloning strategy [24].

C. albicans TLO genes α*1*, α*3*, γ*7*, α*9*, γ*11* and α*12* were heterologously expressed in the *C. dubliniensis* ΔΔ*tlo* double mutant under the control of their native promoters. Each gene was PCR amplified and cloned in pCDRI and introduced in the ΔΔ*tlo* double mutant as described above.

cDNA synthesis and qualitative real-time PCR

RNA was extracted and used to generate cDNA as described by Flanagan *et al*. [26]. qRT-PCR was carried out on the Applied Biosystem 7500 Fast Real Time PCR System as described by Flanagan *et al*. [26]. Plates were set up in triplicate with the endogenous control, *ACT1*, run alongside each target. Results were exported into Microsoft Excel and the delta Ct values calculated for each sample. These were ultimately graphed using GraphPad Prism version 6 (San Diego, California, USA, www.graphpad.com).

Biofilm induction assays

Biofilm mass was determined using crystal violet to quantify biomass. Cells were grown in YEPD at 37˚C overnight with shaking at 200 rpm. Following overnight incubation, 100 ml was removed and transferred to YNB with 100 mM glucose and incubated overnight at 37˚C with shaking at 200 rpm. Following the second night of incubation, cells were washed in 1X PBS and resuspended in 1 ml of YNB with 100 mM glucose at a cell density of 2 x 10^6 cells/ml. 100 μl of each strain was placed in triplicate into a 96 well plate and incubated at 37˚C for 90 min. Following incubation, the medium was aspirated, and cells washed twice with 150 ml 1X PBS. A 100 μl volume of YNB containing 100 mM glucose was placed onto the washed cells and the plates were incubated at 37˚C for 24 and 48 hr. Following incubation, the wells were washed three times with 200 μl sterile 1X PBS to remove non-adherent cells and 110 μl of 0.4% (v/v) crystal violet was added to each well and stained at room temperature for 45 min. The crystal violet was removed and each well washed with 200 μl of dH_2O three times. The wells were de-stained with 200 μl of 95% (v/v) ethanol for 45 min. A 100 μl aliquot of each suspension was transferred to a new 96-well plate and the absorbance measured at an OD_{540} using a Tecan Plate Reader system (Tecan). Results were analysed using GraphPad Prism version 6.

In vivo infection model

Candidal virulence was assessed using the wax moth larva *Galleria mellonella* obtained from Live Foods Direct (Sheffield, England). Larvae were stored at 15˚C in wood shavings in the dark prior to use and those that weighed between 0.20 to 0.30 g were used within 2 weeks of receipt. For each infection experiment, 10 larvae were placed into sterile 9 cm petri dishes lined with Whatman filter paper and wood shavings. *Candida* strains were assigned a random code prior to each experiment to facilitate blind assessment of virulence. For infection, 1 X 10^6 yeast cells in 20 μl were injected into the haemocoel via the last left pro-leg with a 30G insulin U-100 Micro-Fine syringe (BD New Jersey, USA) as described by Cotter *et al*. (2000) [27]. The inoculated larvae were incubated at 30˚C and larval mortality was assessed at 24 h intervals, as described by Cotter *et al*. (2000) [27]. The results were analysed using GraphPad Prism version 6.

Results

Expression of *Candida albicans TLO* genes in wild type *C. dubliniensis*

In order to investigate the effect on phenotype of expanding the repertoire of *TLO* genes in *C. dubliniensis*, we expressed the *C. albicans TLOβ2*, *TLOγ11* and *TLOα12* genes (representing each of the three *CaTLO* clades) in the *C. dubliniensis* WT Wü284 background under the control of a native *TLO* promoter and that of the constitutively expressed *ACT1* gene. In the case of *TLOβ2* for which no promoter sequence was available, we used the *TLOα1* promoter as a proxy native promoter as both *TLOβ2* and *TLOα1* exhibit similar mRNA expression levels [11].

Quantitative Real Time PCR was used to determine the level of expression of each *TLO* under the expression of their native promotor and that of the *ACT1* gene (Fig 1B). The level of

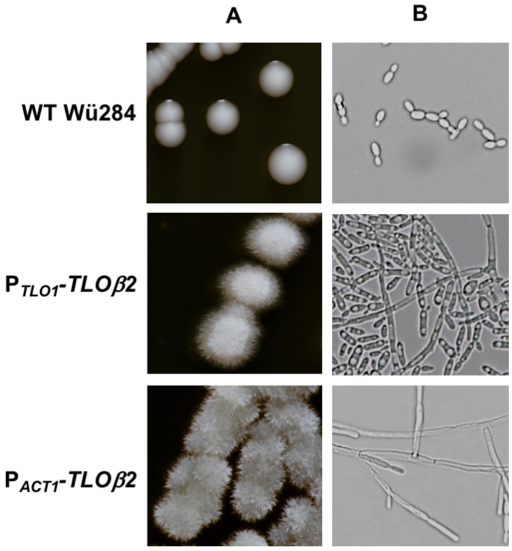


Fig 2. *C. albicans TLO*β2 confers filamentous growth in *C. dubliniensis*. Colony (A) and cellular (B) morphology of *C. dubliniensis* WT Wü284 and derivatives harboring *TLO*β2 expressed from the *TLO1* promoter (P$_{TLO1}$-*TLO*β2) and the *ACT1* promoter (P$_{ACT1}$-*TLO*β2). Colonies were grown for 48 h on solid YEPD agar. The morphology of the cells in representative colonies of each derivative was visualised using a x40 objective lens.

https://doi.org/10.1371/journal.pone.0200852.g002

each *TLO* expressed under their native promotor was lower compared with that of the *ACT1* gene. *TLO*β2, under the expression of the native *TLO*α*1* promoter, was expressed at 0.1 relative to *ACT1*. When placed under the expression of the *ACT1* gene, the expression increased to 1.31 relative to *ACT1*, a fold-change of 13.1. *TLO*γ*11* under the control of its native and the *ACT1* promoters was expressed at 0.45 and 1.05, respectively, relative to *ACT1*, a fold-change of 2.3. Similarly, *TLO*α*12* under the control of the native promoter and *ACT1* gene showed expression levels of 0.005 and 0.5 relative to *ACT1*, a fold-change of 118.

Once the level of expression of each gene had been determined, a range of phenotypic tests was then performed to determine whether the expression of additional *C. albicans TLO* genes had the ability to affect the phenotype of the host strain.

*TLO*β2 expression in wild type *C. dubliniensis* results in hyperfilamentous growth

In wild type *C. dubliniensis* Wü284, expression of *TLO*γ*11* and *TLO*α*12* under the native or *ACT1* promoter did not affect the colony morphology of the strain on YEPD agar. However, *TLO*β*2* whether expressed under the comparatively weak *TLO*α*1* promoter or the *ACT1* promoter in wild-type *C. dubliniensis* resulted in wrinkled colonies on YEPD agar and hypha formation in YEPD broth (Fig 2). This phenotype was affected by expression levels of *TLO*β*2*, with the *TLO*α*1* promoter variant exhibiting a predominantly pseudohyphal mode of growth

in YEPD broth, and the *ACT1* variant producing longer filaments with evidence of true hyphal growth (Fig 2B).

TLO expansion in wild type *C. dubliniensis* affects growth rate

It has previously been shown that deletion of the two *TLO* genes present in the *C. dubliniensis* genome leads to reduced growth rate in YEPD and also results in greatly increased doubling times when galactose is the sole source of carbon [12]. To investigate the ability of an expanded *TLO* gene repertoire to affect growth rate we cultured all strains in YEP-Glucose and YEP-Galactose over an 8 h time course. In WT *C. dubliniensis* Wü284, expression of *TLOα12* under the control of its native or the *ACT1* promoter did not affect doubling times in YEP-Glucose or YEP-Galactose (Fig 3). Expression of *TLOγ11* under its native promoter in WT Wü284 reduced the doubling time in YEP-Glucose by approximately 10 min (Fig 3A). Expression of *TLOβ2* had the effect of greatly reducing growth in both media, and this was most significant in strain P_{ACT1}-*TLOβ2* compared to P_{TLO1}-*TLOβ2* (Fig 3B). This effect on growth rate is most likely due to the filamentous morphology exhibited by these strains (Fig 2).

TLO expansion in wild type *C. dubliniensis* enhances H_2O_2 resistance

Using a broth dilution MIC test for H_2O_2 we showed that expression of TLOγ11 and *TLOα12* in WT *C. dubliniensis* Wü284 under their native promoters led to a doubling of MIC_{80} from 10mM to 20mM. *TLOα12* expressed using the *ACT1* promoter also led to a similar increase in MIC_{80}. Using this assay, expression of *TLOβ2* under the control of the *TLOα1* or *ACT1* promoters did not affect susceptibility to H_2O_2 (Fig 4A).

TLOβ2 increases susceptibility to cell wall damaging agents

In order to determine if the *C. albicans TLO* genes differ in their effects on cell wall stress responses we compared the effect of these genes on growth on media containing the

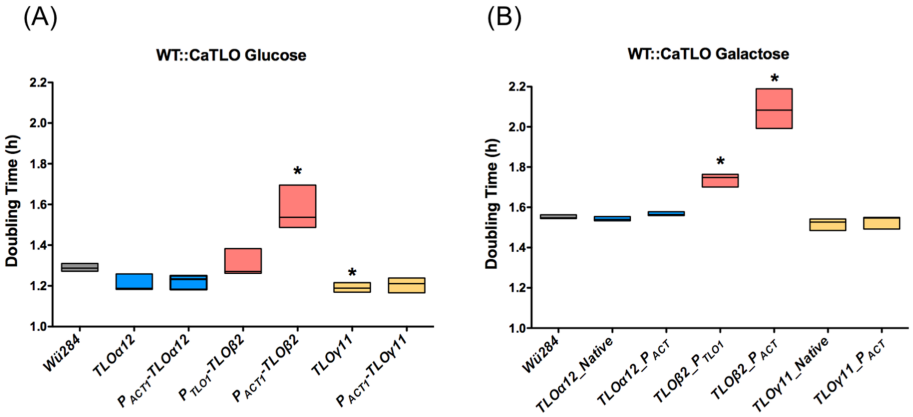


Fig 3. The effect of *C. albicans TLO* genes on growth rates in YEP-Glucose and -Galactose broth. Doubling times of WT Wü284 and derivatives expressing the indicated *C. albicans TLO* genes in YEP-Glucose (A) and -Galactose (B). Asterisks indicate significant difference from Wü284. Data were generated in three replicate experiments.

https://doi.org/10.1371/journal.pone.0200852.g003

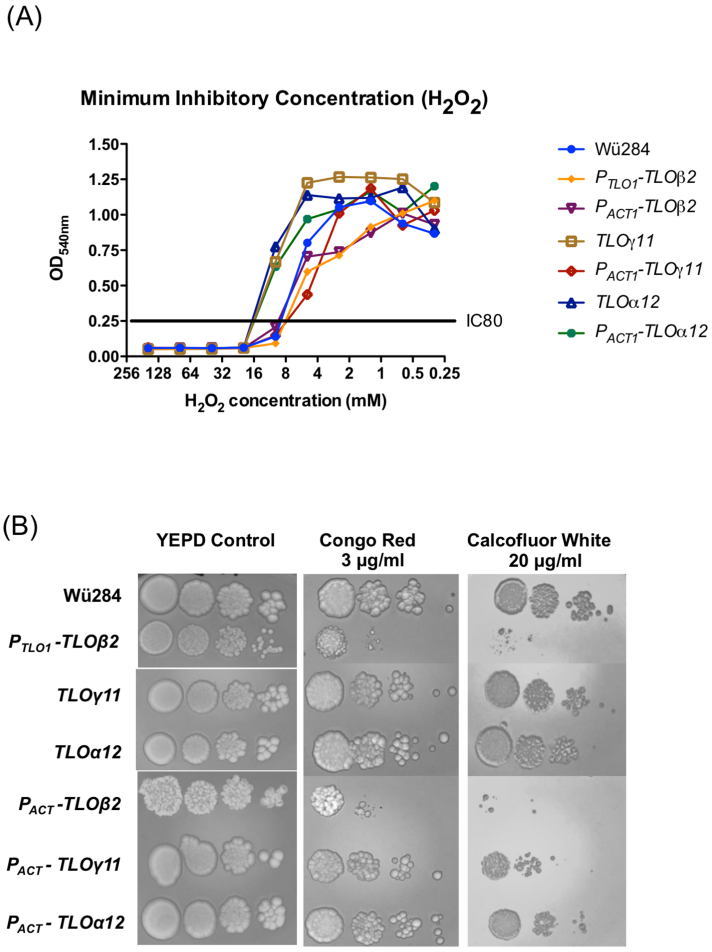


Fig 4. The effect of *C. albicans TLO* genes on susceptibility to H_2O_2 and cell wall damaging agents. (A) Minimum inhibitory concentration of H_2O_2 was determined by broth dilution. The IC_{80} is indicated and shows the concentration of H_2O_2 that reduced growth of the derivatives tested below 80% of the inhibitor-free control. (B) Ten-fold serial dilutions (left to right) of 2 x 10^4 cells were spotted on to plates containing 3 µg/ml Congo Red and 20 µg/ml Calcofluor White.

https://doi.org/10.1371/journal.pone.0200852.g004

β-1,3-glucan-binding dye Congo Red and the chitin-binding dye Calcofluor White. Expression of *TLOγ11 and TLOα12* in WT *C. dubliniensis* Wü284 under their native promoters did not affect susceptibility to Congo Red or Calcofluor White, while expression of *TLO*β*2* under the

control of the *TLOα1* or *ACT1* promoters in wild type *C. dubliniensis* resulted in increased susceptibility to both agents (Fig 4B).

Biofilm formation

The ability to form biofilm on plastic surfaces following 24 h and 48 h incubation was assessed using a Crystal Violet staining assay (Fig 5). Expression of *TLOβ2* in strain Wü284 resulted in a significant decrease in biofilm formation at 24 h (Fig 5). Expression of *TLOγ11* and *TLOα12* resulted in a higher degree of biofilm formation at the 24 h timepoint (Fig 5). The greatest increase in biofilm formation relative to Wü284 was observed in the P_{ACT1}-*TLOγ11* and P_{ACT1}-*TLO α12* expressing strains, which exhibited increased biofilm at 24 h and 48 h (Fig 5).

TLO expansion in wild type *C. dubliniensis* enhances virulence in the *Galleria mellonella* infection model

Given the differential effects of specific *TLO* genes on various virulence attributes, such as morphology, stress tolerance and cell wall integrity, we decided to investigate if differences in virulence could be detected using an *in vivo* infection model. Virulence of WT *C. dubliniensis* Wü284 expressing *TLOβ2*, *TLOγ11* and *TLOα12* from a native *TLO* promoter or the *ACT1* promoter was investigated using the insect larval *G. mellonella* model. Although genes expressed from native *TLO* promoters did not confer significant increases in virulence, expression of these *C. albicans TLO* genes in WT *C. dubliniensis* under the control of the *ACT1* promoter significantly enhanced virulence in this larval infection model (Fig 6). This effect was most significant in the case of the P_{ACT1}- *TLOβ2* expressing strain.

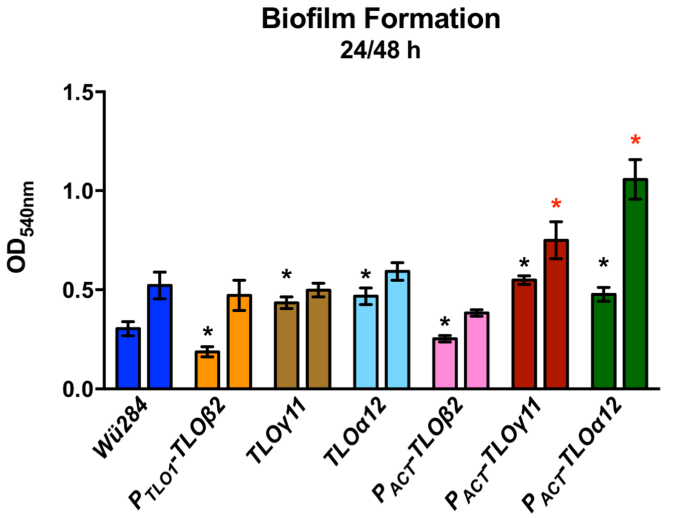


Fig 5. The effect of *C. albicans TLO* genes on biofilm formation on plastic surfaces. Each strain was grown in Spider medium in a 96-well plate for 48 h and biomass measured using a Crystal Violet assay. An asterisk indicates significant differences with * at 24 h and red * at 48 h.

https://doi.org/10.1371/journal.pone.0200852.g005

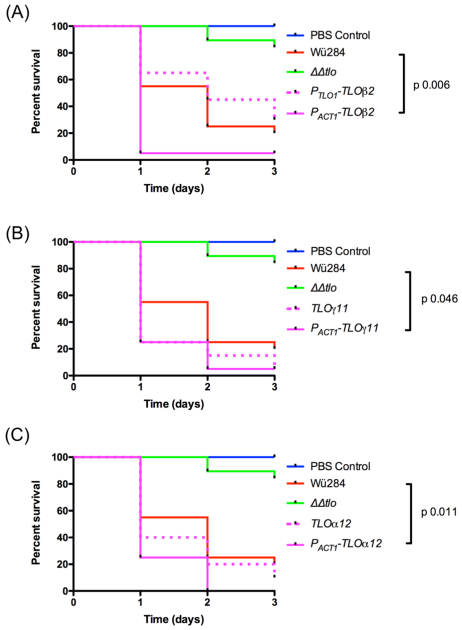


Fig 6. The effect of *C. albicans TLO* genes on the virulence of wild-type Wü284. Ten *G. mellonella* larvae were inoculated with 1 X 10^6 cells of each indicated strain (performed blind) and viability was monitored over 3 days. Results presented represent three independent infection experiments. *P* values indicate results of a Log-Rank (Mantel-Cox) test against the wild type Wü284 survival curve.

https://doi.org/10.1371/journal.pone.0200852.g006

Expression of *Candida albicans TLO* genes in the *C. dubliniensis Δtlo1/Δtlo2* (*ΔΔtlo*) double mutant

In order to further investigate the range of phenotypes regulated by *C. albicans* Tlo proteins, we also expressed representative members of the *C. albicans TLO* gene family in the *C. dubliniensis Δtlo1/Δtlo2* (*ΔΔtlo*) double mutant under the control of their native upstream regulatory elements. Several attempts were made to generate stable transformants expressing *TLO*β2, however, no viable transformants were recovered in these experiments. Quantitative Real Time PCR was used to determine the level of expression of the *C. albicans TLO*s in the *C. dubliniensis ΔΔtlo* backgrounds (S1 Fig). *TLO 1*, *3*, *9* and *12*, all of which belong to the αclade, show similar expression levels of expression relative to *ACT1* (0.016 to 0.088, Fig 1B). Interestingly, the γ clade genes *TLOγ7* (0.38 relative to *ACT1)* and *TLOγ11* (0.006 relative to *ACT1*) differed greatly in expression levels compared with one another (S1 Fig).

All *TLO* genes tested, with the exception of *TLOγ7*, restored filamentous growth in the *C. dubliniensis ΔΔtlo* mutant, which is normally not capable of forming true hyphae (Fig 7A). *TLOα9*, *γ11* and *α12* restored the ability to produce chlamydospores in the deletion mutant, while *TLOα1*, *α3* and *γ7* were unable to do so (Fig 7B).

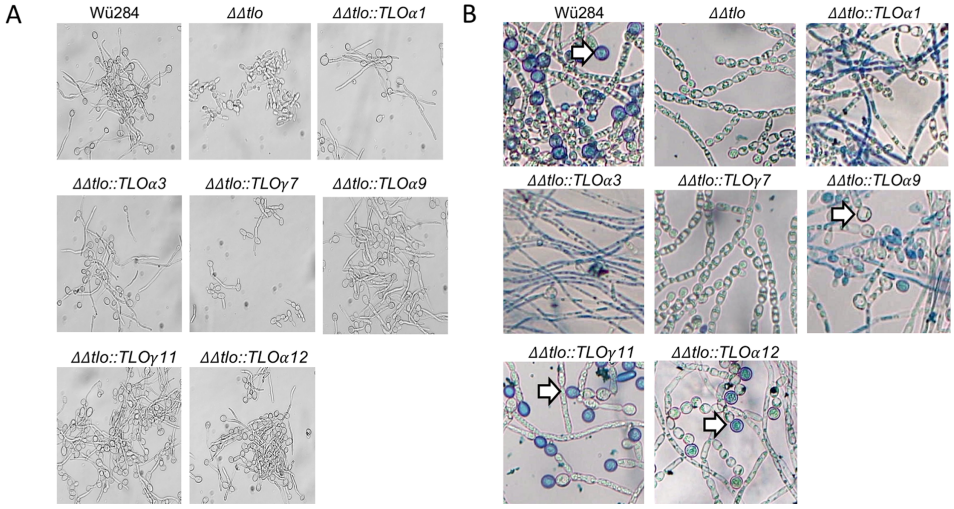


Fig 7. Morphology of *C. dubliniensis* ΔΔ*tlo* expressing *CaTLO* genes. A. Photomicrographs of *C. dubliniensis* ΔΔ*tlo* and derivatives harboring the indicated *C. albicans TLO* genes following 4 h growth in water supplemented with 10% (v/v) foetal bovine serum at 37˚C. B. Chlamydospore formation of *C. dubliniensis* ΔΔ*tlo* and derivatives harboring the indicated *C. albicans TLO* genes on cornmeal agar supplemented with tween. Chlamydospores are indicated by arrows. Identical results were observed in replicate experiments.

https://doi.org/10.1371/journal.pone.0200852.g007

All of the *TLO* genes complemented the defective growth of the ***ΔΔtlo*** mutant in YEP-Galactose (Fig 8A and 8B). The ***ΔΔtlo*** *C. dubliniensis* mutant has previously been shown to produce excess levels of biofilm on plastic surfaces relative to wild type [12]. Following 24 h growth under biofilm forming conditions, the ***ΔΔtlo***::*TLOα1*, *α3*, and *α9* strains exhibited reduced biofilm formation relative to the ***ΔΔtlo*** double mutant and comparable to that observed with WT Wü284. The remaining genes tested either resulted in similar or greater (e.g. *TLOγ7*) levels of biofilm than the ***ΔΔtlo*** double mutant (Fig 8C).

TLOγ11 also conferred increased resistance to oxidative stress. At a concentration of 6 mM H_2O_2 ***ΔΔtlo***::*TLOγ11* had the greatest effect on enhancing tolerance of oxidative stress, with the remaining genes conferring tolerance, but to a lesser extent (Fig 9A). When incubated on solid YEPD supplemented with 1 M NaCl, all *CaTLO* genes tested, with the exception *CaTLOγ11*, resulted in increased growth compared to the ***ΔΔtlo*** mutant strain (Fig 9B). In the presence of the cell wall perturbing compounds Congo Red (2 μg/ml) and Calcofluor White (10 μg/ml) *TLOα3* consistently restored growth of the ***ΔΔtlo*** mutant to wild-type levels. ***ΔΔtlo***::*TLOα12* also exhibited enhanced levels of growth on Calcofluor White (10 μg/ml) compared to the ***ΔΔtlo*** mutant (Fig 9C and 9D).

Finally, examination of the virulence of the *CaTLO* expressing strains in the *Galleria mellonella* model showed that ***ΔΔtlo***::*TLOα3* enhanced virulence to the greatest extent, with survival rates significantly less than the parental ***ΔΔtlo*** mutant and similar to the WT strain (Fig 10). The remaining *TLO* genes tested were shown to result in a restoration of virulence in the infection model with mortality rates greater than that of the ***ΔΔtlo*** double mutant but less than the WT Wü284.

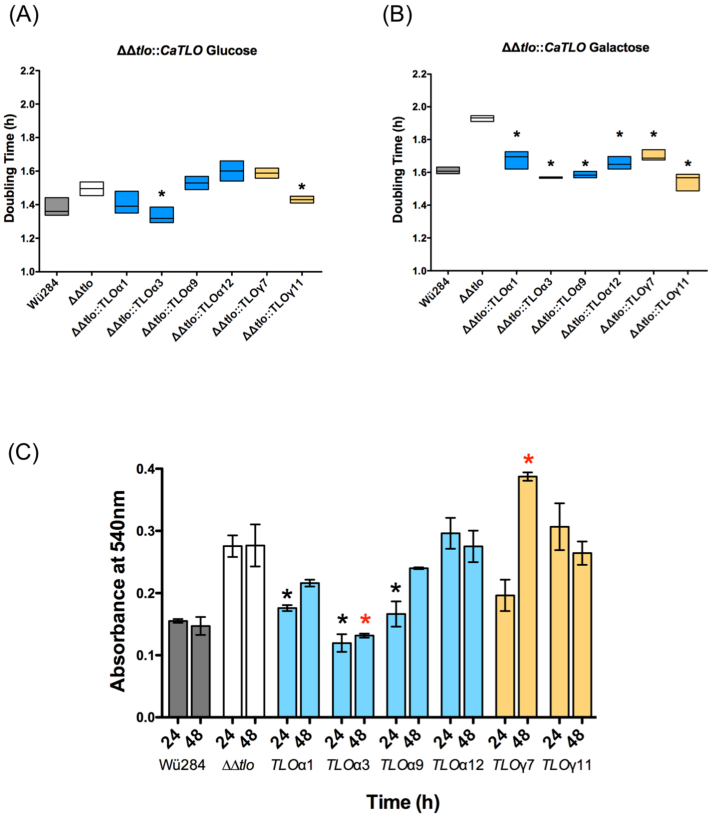


Fig 8. Growth of *C. dubliniensis* ΔΔ*tlo* and derivatives harboring *C. albicans TLO* genes. A and B show doubling times of WT Wü284, the ΔΔ*tlo* double mutant and derivatives expressing indicated *C. albicans TLO* genes in YEP-Glucose (A) and -Galactose (B). Stars indicate strains exhibiting doubling times significantly different from ΔΔ*tlo* ($p \leq 0.05$). Panel C shows biofilm formation on plastic surfaces. Each ΔΔ*tlo*::*TLO* strain was grown in the presence of YEPD in a 96-well plate for 48 h. Biomass was measured using a crystal violet assay in three replicate experiments. Asterisks indicate significant differences from ΔΔ*tlo* at 24 h (*) and 48 h (red *), respectively. Data are the result of three independent experiments.

https://doi.org/10.1371/journal.pone.0200852.g008

Discussion

One of the largest gene families in *C. albicans* is the *TLO* family, which consists of up to 15 members, each encoding a protein orthologous to the Med2 subunit of the transcriptional regulator complex Mediator [28]. This expansion is unique to *C. albicans* and there is significant variation in the copy number of genes in this family between different strains [14]. There are only two *TLO* genes encoded in the genome of *C. dubliniensis*, the species most closely related

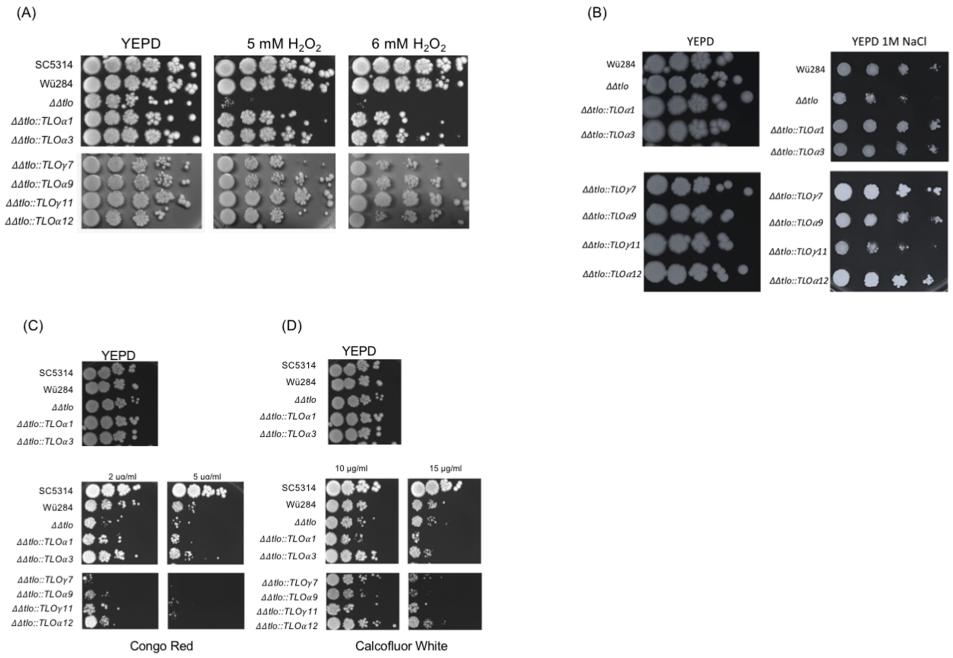


Fig 9. *C. albicans TLO* genes differentially affect tolerance of environmental stress conditions. Growth of each ***ΔΔtlo*::*CaTLO*** strain in the presence (A) H_2O_2, (B) NaCl, (C) Congo Red and (D) Calcofluor White. Ten-fold serial dilutions (left to right) of 2 x 10^4 cells were spotted onto YEPD agar and YEPD agar containing the indicated agents. Plates were incubated for 48 h at 37˚C.

https://doi.org/10.1371/journal.pone.0200852.g009

species to *C. albicans*. Deletion of the two *TLO* genes in *C. dubliniensis* resulted in significant transcriptional and phenotypic defects, including an inability to produce hyphae and reduced tolerance of oxidative stress. Reintroduction of each of the *C. dubliniensis TLO* genes into the double mutant background indicated that the *CdTLO1* and *CdTLO2* genes differ in their ability to complement the mutant phenotypes, suggesting they may have distinct functions in gene control [12]. The purpose of the current study was to investigate if expansion of this two-membered family in *C. dubliniensis* affects phenotypes associated with virulence. Such a finding would lend support to our hypothesis that expansion of the *C. albicans TLO* gene family played a role in the evolution of the enhanced virulence of this species in comparison with other related species.

We expanded the repertoire of Tlo proteins in *C. dubliniensis* Wü284 using representatives of the α, β, and γ *TLO* families, namely *TLOβ2*, *TLOγ11* and *TLOα12* and a summary of the effects of this expansion is shown in Fig 11A. Expression of *TLOβ2* had the most dramatic effect on morphology, resulting in the production of wrinkled colonies containing cells with hyphal morphologies. Interestingly, the extent of this phenotype was influenced by the expression level of *TLOβ2*, with the *ACT1*-promoter driven gene resulting in more highly-wrinkled

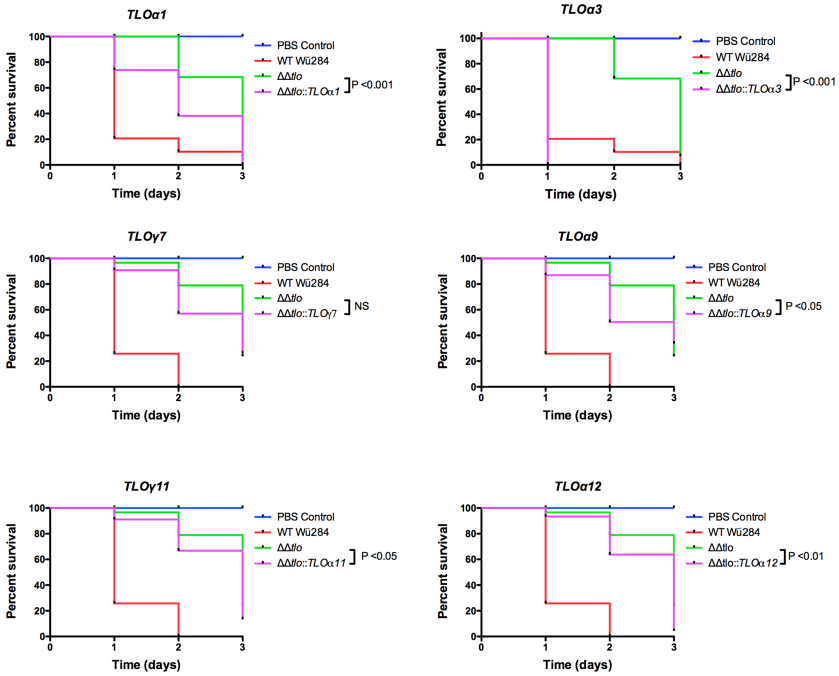


Fig 10. The effect of *C. albicans TLO* genes on the virulence of the *ΔΔtlo* mutant. Ten *G. mellonella larvae* were inoculated with 1 X 10^6 cells of each indicated strain (performed blind) and viability was monitored over 3 days. Results presented represent three independent infection experiments. *P* values indicate results of a Log-Rank (Mantel-Cox) test against the ***ΔΔtlo*** mutant survival curve.

https://doi.org/10.1371/journal.pone.0200852.g010

colonies with a higher proportion of true hyphae. The more pronounced phenotype in the highly expressed construct indicates that expression level influences phenotype. This affect may be exerted by displacing endogenous CdTlo1 and CdTlo2 from the Mediator complex, therefore promoting Tloβ2 regulated functions. Alternatively, the higher expression levels may create a pool of Tlo in excess of Mediator. A similar phenotype was recently described following overexpression of CdTlo2 to create a Mediator excess population of Tlo in *C. dubliniensis* [20]. Unexpectedly, *TLOβ2* could not be expressed in the *C. dubliniensis* Δ*Δtlo* double mutant, suggesting that a Mediator complex exclusively containing Tloβ2 is lethal to the cell.

Although wild-type Wü284 expressing *TLOγ11* and *TLOα12* were not affected in morphology, these strains exhibited a specific enhanced resistance to H_2O_2, whereas the *TLOβ2* expressing strains which were filamentous were highly susceptible to oxidative and cell wall stress, suggesting that the morphology of the cells may contribute to their ability to tolerate stress. Expression of *TLOβ2*, *TLOγ11* and *TLOα12* in wild type Wü284 had varying effects on growth rates *in vitro*. *TLOβ2* generally increased doubling times in YEP-Glucose and -Galactose and this might be related to the hyper-filamentous, polarised growth pattern exhibited by

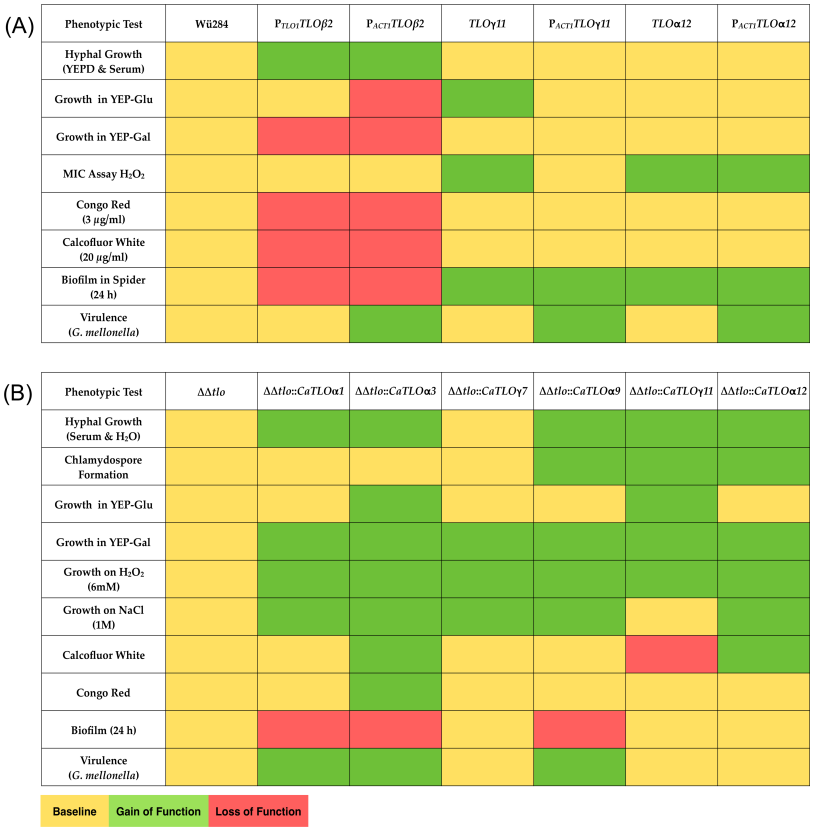


Fig 11. Heat maps summarizing phenotypic effects of *CaTLO* genes. The phenotypic effects of expressing each *CaTLO* gene in the *C. dubliniensis* wild-type (panel A) and in *C. dubliniensis* ΔΔ*tlo* (panel B) are colour coded; yellow indicates the same phenotype as the mutant, green indicates a gain of function and red indicates a loss of function.

https://doi.org/10.1371/journal.pone.0200852.g011

this strain. In contrast, *TLOγ11* transformants exhibited a small (approximately 10 min) but significant reduction in doubling time in YEP-Glucose.

Infection of *G. mellonella* larvae with these strains showed that expression of the heterologous genes at low level using native *TLO* promoters had limited effects on larval survival, however expression under the control of the stronger *ACT1* promoter resulted in significantly reduced larval survival (similar to *C. albicans* SC5314). These data for the first time provide experimental evidence supporting enhanced fitness as a result of *TLO* copy number expansion in *Candida species*. It is interesting to note that the phenotypic effect was most significant

when the genes were expressed under the control of the stronger *ACT1* promoter, indicating that a critical level of Tlo is required for this gain of function. This may partly explain why *C. albicans* has expanded the *TLO* family to such a significant extent.

In the second part of our study, to better understand the diversity of functions regulated by *TLO*s, we expressed a range of *C. albicans TLO* genes in a ΔΔ*tlo C. dubliniensis* background. In general, heterologous expression of the *C. albicans TLO* genes in the ΔΔ*tlo C. dubliniensis* background could restore all phenotypes examined, including some subtle and some major differences in the ability of individual paralogs to complement the phenotypes. A summary of these phenoptypes is shown in Fig 11B. It is clear from this heatmap that individual *CaTLO*s differ in their ability to affect specific phenotypes and in the magnitude of their restorative capability. For example, *TLOα1*, *α3* and *γ7* do not have the ability to restore chlamydospore production in the mutant, while *TLOγ11* and *α12* did not suppress biofilm formation (despite being active inducers of stress responses). The exception to this was *TLOγ7*, which despite having the highest expression level of all of the *C. albicans TLO*s (0.38 relative to *ACT1*) tested in the ΔΔ*tlo C. dubliniensis* background, had the least effect on restoring the phenotypes in the mutant. *TLOγ7* failed to restore the ability to form hyphae and chlamydospores or growth in media containing cell perturbing compounds. However, *TLOγ7* did restore growth in YEP-Gal and tolerance of sodium chloride and H_2O_2, indicating that the gene possesses some functionality. In order to investigate if there are differences in the effects of specific *CaTLO* genes on virulence we tested the virulence of strains in the *Galleria mellonella* larval infection model. *TLOα3* was found to restore virulence in the *C. dubliniensis* ΔΔ*tlo* mutant to a greater extent than the other *TLO* genes tested, suggesting a clear disparity in the activity of specific *TLO* genes.

These data complement the findings of Dunn *et al*. [21] who used a 'Tet-ON' missexpression system to probe individual *TLO* genes in *C. albicans* SC5314. The authors concluded that *TLO*s controlled multiple phenotypes and that single phenotypes were often regulated by multiple *TLO*s, including virulence in *G. mellonella*. However, it is difficult to directly compare the results of individual phenotypic tests in both studies due the different nature of the host strains (*C. albicans* and *C. dubliniensis*) and the phenotypic tests used.

In summary, the *C. albicans TLO* gene family is comprised of fifteen genes, mainly situated in the subtelomeric region of the chromosomes. These regions have been shown in other organisms to undergo rapid evolution and gene families in the subtelomeres have been demonstrated to expand rapidly and undergo functional divergence [29]. Our data support the hypothesis that there is functional diversity in the *C. albicans TLO* gene family and also indicate that the high copy number of *TLO* genes in C. *albicans* may have evolved to increase gene dosage, which in our larval infection model has a significant effect on virulence. Studies are now underway to confirm these hypotheses by attempting to deplete the *TLO* gene family in *C. albicans* using CRISPR-Cas9 mutagenesis.

Supporting information

S1 Fig. Expression of *CaTLO* genes in *C. dubliniensis* the ΔΔ*tlo* mutant. RT-PCR expression data of respresentative *C. albicans TLO* genes expressed in the *C. dubliniensis* ΔΔ*tlo* mutant strain. RT-PCR expression graphs represent three independent experiments.
(TIF)

S1 Table. List of strains used in this study.
(DOCX)

S2 Table. Sequence of oligonucleotides used in this study.
(DOCX)

S1 File. DNA sequence of P_{TLO1}-*TLOβ2* gene fusion.
(TXT)

Author Contributions

Conceptualization: Gary P. Moran, Derek J. Sullivan.

Formal analysis: Peter R. Flanagan, Gary P. Moran.

Investigation: Peter R. Flanagan, Jessica Fletcher, Hannah Boyle, Razvan Sulea, Gary P. Moran.

Project administration: Derek J. Sullivan.

Supervision: Gary P. Moran, Derek J. Sullivan.

Writing – original draft: Peter R. Flanagan.

Writing – review & editing: Peter R. Flanagan, Jessica Fletcher, Gary P. Moran, Derek J. Sullivan.

References

1. Yang YL. Virulence factors of *Candida species*. J Microbiol Immunol Infect. 2003; 36: 223–228. PMID: 14723249
2. Mendes Giannini MJS, Bernardi T, Scorzoni L, Fusco-Almeida AM, Sardi JCO. *Candida* species: current epidemiology, pathogenicity, biofilm formation, natural antifungal products and new therapeutic options. J Med Microbiol. 2013; 62:10–24. https://doi.org/10.1099/jmm.0.045054-0 PMID: 23180477
3. Hube B. From commensal to pathogen: stage- and tissue-specific gene expression of *Candida albicans*. Curr Opin Microbiol. 2004; 7: 336–341. https://doi.org/10.1016/j.mib.2004.06.003 PMID: 15288621
4. Gow NAR, Brown AJP, Odds FC. Fungal morphogenesis and host invasion. Curr Opin Microbiol. 2002; 5: 366–371. PMID: 12160854
5. Heilmann CJ, Sorgo AG, Siliakus AR, Dekker HL, Brul S, De Koster CG, et al. Hyphal induction in the human fungal pathogen *Candida albicans* reveals a characteristic wall protein profile. Microbiology. 2011; 157: 2297–2307. https://doi.org/10.1099/mic.0.049395-0 PMID: 21602216
6. Lu Y, Su C, Liu H. *Candida albicans* hyphal initiation and elongation. Trends Microbiol. 2014; 22: 707–714. https://doi.org/10.1016/j.tim.2014.09.001 PMID: 25262420
7. Nobile CJ, Johnson AD. *Candida albicans* biofilms and human disease. Annu Rev Microbiol. 2015; 69: 71–92. https://doi.org/10.1146/annurev-micro-091014-104330 PMID: 26488273
8. Sullivan DJ, Westerneng TJ, Haynes KA, Bennett DE, Coleman DC. *Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. Microbiology. 1995; 141: 1507–1521. https://doi.org/10.1099/13500872-141-7-1507 PMID: 7551019
9. Jackson AP, Gamble JA, Yeomans T, Moran GP, Saunders D, Harris D, et al. Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. Genome Res. 2009; 19: 2231–2244. https://doi.org/10.1101/gr.097501.109 PMID: 19745113
10. Moran GP, Coleman DC, Sullivan DJ. *Candida albicans* versus *Candida dubliniensis*: Why Is *C. albicans* more pathogenic? Int J Microbiol. 2012; 2012: 205921–205927. https://doi.org/10.1155/2012/205921 PMID: 21904553
11. Anderson MZ, Baller JA, Dulmage K, Wigen L, Berman J. The three clades of the telomere-associated TLO gene family of *Candida albicans* have different splicing, localization, and expression features. Eukaryotic Cell. American Society for Microbiology; 2012; 11: 1268–1275.
12. Haran J, Boyle H, Hokamp K, Yeomans T, Liu Z, Church M, et al. Telomeric ORFs (TLOs) in *Candida* spp. encode mediator subunits that regulate distinct virulence traits. PLoS Genet. 2014; 10: e1004658. https://doi.org/10.1371/journal.pgen.1004658 PMID: 25356803
13. Anderson MZ, Wigen LJ, Burrack LS, Berman J. Real-Time evolution of a subtelomeric gene family in *Candida albicans*. Genetics; 2015; 200: 907–919. https://doi.org/10.1534/genetics.115.177451 PMID: 25956943

14. Hirakawa MP, Martinez DA, Sakthikumar S, Anderson MZ, Berlin A, Gujja S, et al. Genetic and phenotypic intra-species variation in *Candida albicans*. Genome Res. 2015; 25: 413–425. https://doi.org/10.1101/gr.174623.114 PMID: 25504520

15. Zhang A, Petrov KO, Hyun ER, Liu Z, Gerber SA, Myers LC. The Tlo proteins are stoichiometric components of *Candida albicans* mediator anchored via the Med3 subunit. Eukaryot Cell. 2012; 11: 874–884. https://doi.org/10.1128/EC.00095-12 PMID: 22562472

16. Conaway RC, Conaway JW. Function and regulation of the mediator complex. Curr Opin Genet Devel. 2011; 21: 225–230.

17. Uwamahoro N, Qu Y, Jelicic B, Lo TL, Beaurepaire C, Bantun F, et al. The functions of mediator in *Candida albicans* support a role in shaping species-specific gene expression. PLoS Genet. 2012; 8: e1002613. https://doi.org/10.1371/journal.pgen.1002613 PMID: 22496666

18. Nishikawa JL, Boeszoermenyi A, Vale-Silva LA, Torelli R, Posteraro B, Sohn Y- J, et al. Inhibiting fungal multidrug resistance by disrupting an activator-mediator interaction. Nature. 2016; 530: 485–489. https://doi.org/10.1038/nature16963 PMID: 26886795

19. Liu Z, Myers LC. Mediator tail module is required for Tac1-activated *CDR1* expression and azole resistance in *Candida albicans*. Antimicrob Agents Chemother. 2017; 61: e01342–17. https://doi.org/10.1128/AAC.01342-17 PMID: 28807920

20. Liu Z, Moran GP, Sullivan DJ, Maccallum DM, Myers LC. Amplification of TLO mediator subunit genes facilitate filamentous growth in *Candida* spp. PLoS Genet. 2016; 12: e1006373. https://doi.org/10.1371/journal.pgen.1006373 PMID: 27741243

21. Dunn MJ, Kinney GM, Washington PM, Berman J, Anderson MZ. Functional diversification accompanies gene family expansion of MED2 homologs in *Candida albicans*. PLoS Genet. 2018; 14: e1007326. https://doi.org/10.1371/journal.pgen.1007326 PMID: 29630599

22. Lee KL, Buckley HR, Campbell CC. An amino acid liquid synthetic medium for the development of mycellal and yeast forms of *Candida albicans*. Sabouraudia: J Med Vet Mycol. 2009; 13: 148–153

23. Liu H, Köhler J, Fink GR. Suppression of hyphal formation in Candida albicans by mutation of a STE12 homolog. Science. 1994; 266: 1723–1726. PMID: 7992058

24. Moran GP, Maccallum DM, Spiering MJ, Coleman DC, Sullivan DJ. Differential regulation of the transcriptional repressor *NRG1* accounts for altered host-cell interactions in *Candida albicans* and *Candida dubliniensis*. Mol Microbiol. 2007; 66: 915–929. https://doi.org/10.1111/j.1365-2958.2007.05965.x PMID: 17927699

25. Staib P, Moran GP, Sullivan DJ, Coleman DC, Morschhauser J. Isogenic strain construction and gene targeting in *Candida dubliniensis*. J Bacteriol. 2001; 183: 2859–2865. https://doi.org/10.1128/JB.183.9.2859-2865.2001 PMID: 11292806

26. Flanagan PR, Liu N-N, Fitzpatrick DJ, Hokamp K, Köhler JR, Moran GP, et al. The *Candida albicans* TOR-activating GTPases Gtr1 and Rhb1 coregulate starvation responses and biofilm formation. mSphere. 2017; 2: e00477–17. https://doi.org/10.1128/mSphere.00477-17 PMID: 29152581

27. Cotter G, Doyle S, Kavanagh K. Development of an insect model for the *in vivo* pathogenicity testing of yeasts. FEMS Immunol Med Microbiol. 2000; 27: 163–169. PMID: 10640612

28. Van Het Hoog M, Rast TJ, Martchenko M, Grindle S, Dignard D, Hogues H, et al. Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. Genome Biol. 2007; 8: R52. https://doi.org/10.1186/gb-2007-8-4-r52 PMID: 17419877

29. Brown CA, Murray AW, Verstrepen KJ. Rapid expansion and functional divergence of subtelomeric gene families in yeasts. Curr Biol. Elsevier; 2010; 20: 895–903. https://doi.org/10.1016/j.cub.2010.04.027 PMID: 20471265

Contribution to publication: Investigation and data generation for Wü284 WT + *CaTLO* strains, as well as reviwing and editing the draft manuscript.

Chapter 1

General Introduction

1.1 *Candida* species and disease

Fungal infections are a serious cause of mortality worldwide, with fungal diseases killing 1.5 million people annually. Most of these deaths are completely preventable, as many fungal diseases are complications of other health problems (Bongomin *et al.*, 2017). While there are over 1.5 million different species of fungi, only about 300 are known to cause disease in humans and of these *Candida* spp. are the cause of the majority of fungal diseases (Bartkowiak and Greenleaf, 2011).

Candida species are budding yeasts and are members of the Ascomycota phylum of fungi, and the sub-phylum Saccharomycotina. This sub-phylum contains one class Saccharomycetes, which in turn contains one order Saccharomycetales, and it is in this order where the *Candida* genus can be found (McManus and Coleman, 2014). The majority of clinically significant *Candida* species are included in the CTG clade, which includes species that encode the codon CTG as serine rather than leucine (Turner and Butler, 2014). The whole genome duplication (WGD) clade is another grouping in the Saccharomycotina sub-phylum, and this clade contains *C. glabrata* and *C. krusei*. A phylogenetic tree showing *Candida* spp. relatedness can be seen in Figure 1.1, where it is also shown that *Candida* species are related to *Saccharomyces cerevisiae*, which is a member of the WGD clade.

Studies on the oral microbiome have mostly focused on bacterial populations, but fungi, protozoa, viruses and archaea can all be found in the oral cavity (Deo and Deshmukh, 2019). *Candida* species are the most common species of fungi isolated from the oral microbiome of healthy people, and it has recently been estimated that up to 80% of humans carry some species of *Candida* as part of their normal oral microbial flora, and over around 50% carry *C. albicans* specifically (Talapko *et al.*, 2021). *Candida* spp. can colonize the oral cavity, the skin, the gastrointestinal tract, and the vaginal canal. Often oral colonization of newborns occurs during vaginal childbirth (Talapko *et al.*, 2021, Al-Rusan *et al.*, 2017).

However, when dysbiosis occurs in the microbiome, *Candida* species can overgrow and invade the epithelium, resulting in an infection (Willems *et al.*, 2020). There are estimated to be 17 species of *Candida* that can cause disease in humans, but of these, five species cause 90% of infections. They are *C. albicans*, *C. glabrata*, *C. parapsilosis*, *C. tropicalis* and *C. krusei* (Sardi *et al.*, 2013). Another species, *C. dubliniensis,* which is more

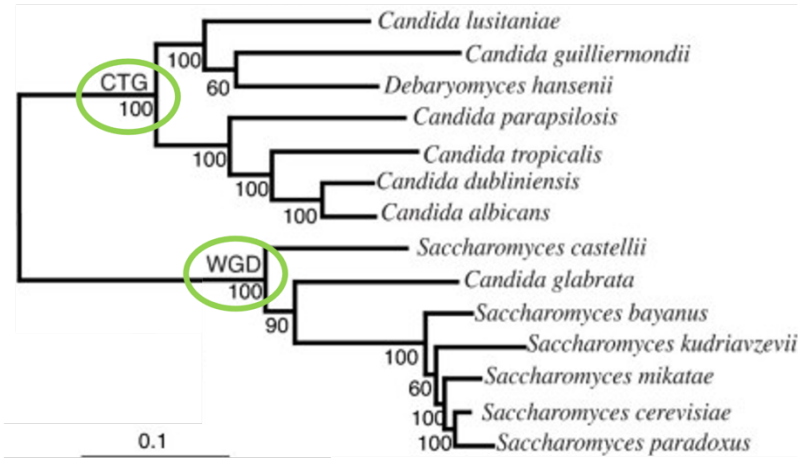


Figure 1.1 Phylogenetic tree showing relatedness of *Candida* species

Maximum likelihood phylogenetic tree adapted from (Fitzpatrick *et al.*, 2006) based on alignment of 153 conserved fungal genes. The numbers at the base of each node indicate the percentage confidence in the node, and the scale bar indicates the distance on the graph that is equal to 0.1 nucleotide substitutions per site (number of substitutions divided by length of sequence). The CTG clade and the WGD are distinct, the emergence of the two is highlighted with green circles on the tree. All pathogenic *Candida* strains apart from *C. glabrata* and *C. krusei* fall into the CTG clade. *C. glabrata* can be seen on the tree in the WGD clade, *C. krusei* is not shown.

closely related to *C. albicans* than to any other *Candida* species, and which is less commonly found in the human microbiota and in infection, will also be discussed in this work. While in recent years there has been somewhat of a shift, with non-*albicans* species causing more and more infections, most epidemiological studies suggest that *C. albicans* is still the most common causative species in human *Candida* infection, causing up to 85% of global *Candida* infections (Goncalves *et al.*, 2016). Non-*albicans* species, such as *C. glabrata* and *C. kruseii*, are generally more resistant to the azole class of antifungals than *C. albicans*, and their recent rise in prominence is thought to be the effect of increased use of azoles to treat *Candida* infections (Goncalves *et al.*, 2016). *Candida* spp. can cause a wide range of infections, from mild to severe. A summary of the main presentations of *Candida* infections can be seen in Figure 1.2 (A), where estimated annual case numbers can also be found. The term 'candidiasis' is used to group infections caused by *Candida* spp., these can be mucosal, cutaneous or more deep-seeded infections (Pappas *et al.*, 2018). Though the causative *Candida* species can differ, clinical presentation of candidiasis is the same no matter the species (Vila *et al.*, 2020). Often superficial infections of the mucosae (e.g. the vagina and oral cavity) can be cleared following treatment, but in certain cases, a patient can be plagued with recurrent *Candida* infections. Of pathogenic fungi, *C. albicans* is the main fungus responsible for mucosal disease (Bongomin *et al.*, 2017).

1.1.1 Oral candidiasis

Oral candidiasis (OC) is a *Candida* infection of the oral cavity and is thought to affect up to 2 million people annually. A further complication of OC is the invasion of the epithelial lining of the oropharynx, termed oropharyngeal candidiasis (OPC), and which is thought to affect up to 1.3 million people annually (Bongomin *et al.*, 2017, Vila *et al.*, 2020). The most common form of OC is pseudomembranous candidiasis, which is commonly referred to as 'thrush'. This presents as white/yellow plaques throughout the oral cavity that can be scraped away easily, as seen in Figure 1.2 (B), (Vila *et al.*, 2020). This infection normally occurs in neonates (who may have picked up *Candida* species as they travelled through the vaginal canal during birth), those who are immunodeficient, those using topical or inhaled steroids and those who experience chronic dry mouth (Talapko *et al.*, 2021). In HIV patients, OPC is the most common recurring infection, occurring in around 80-90% of all HIV patients. The prevalence of OPC in the HIV community has decreased in the years following the introduction of anti-retroviral therapy, due to the increased immune function of patients receiving this treatment (Patil *et al.*, 2018). Cases

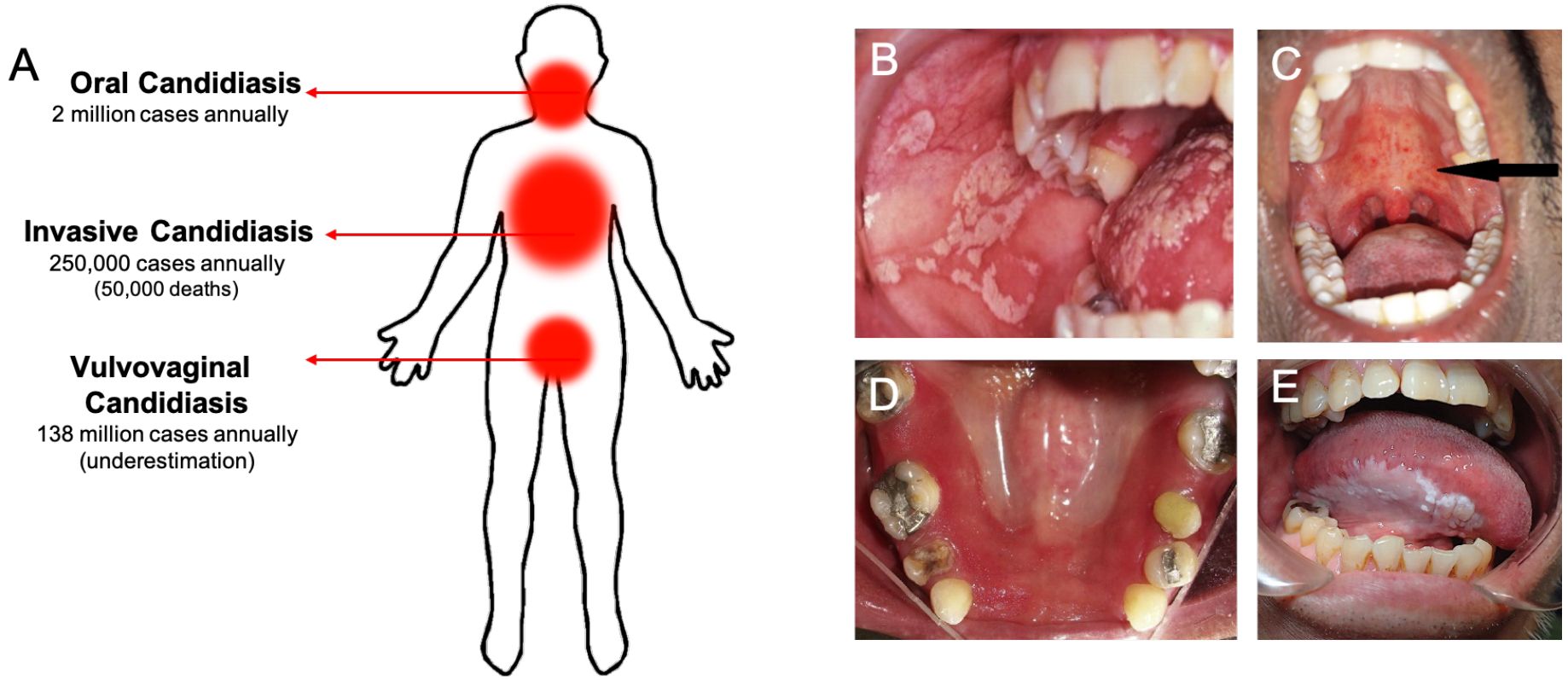


Figure 1.2 *Candida* infections in the human host

(A) An overview of the main sites and types of *Candida* infection in the human host, accompanied by the number of cases of each annually. (B) Pseudomembranous candidiasis on the tongue and buccal mucosa (Jabra-Rizk *et al.*, 2016). (C) Erythematous candidiasis in the oral cavity of a 23 year old, HIV-positive male (Bakshi, 2018). (D) Denture stomatitis associated with the wearing of a removable maxillary partial acrylic denture (Muzyka and Epifanio, 2013). (E) Leukoplakia on the right lateral border of the tongue of a 70 year old man (Ohta and Yoshimura, 2020).

of candidiasis caused by less common species of *Candida* are often seen in patients who are already immunocompromised in some way. In the healthy oral microbiome, *C. dubliniensis* is a very rare constituent, however, in HIV-positive patients, there is an increase in prevalence of *C. dubliniensis* in the oral cavity in the non-disease state. *C. dubliniensis* is not commonly seen to be the cause of OC in otherwise healthy patients, nor is it commonly detected in the oral microbiota, however, in patients who are HIV-positive presenting with OC, *C. dubliniensis* is found at a much higher rate, likely due to their predisposition to be colonised by it (Loreto *et al.*, 2010).

As mentioned previously, *Candida* species are a major component of the human mycobiota, particularly in the oral cavity. When antibiotic treatment causes dysbiosis in the oral cavity, *Candida* can overgrow and temporarily cause acute erythematous candidiasis, characterised by painful red lesions in the mouth, which can be seen in Figure 1.2 (C). Erythematous candidiasis can also affect people who are HIV-positive. In cases caused by antibiotic therapy, when microbial balance is restored after the treatment is finished, the infection is normally cleared without any need for additional treatment (Vila *et al.*, 2020).

Denture stomatitis is a chronic *Candida* infection in the oral cavity, localised to the mucosa bearing a denture or prosthesis, often occurring in people who are otherwise healthy individuals, see Figure 1.2 (D), (Jabra-Rizk *et al.*, 2016). Denture stomatitis affects up to 70% of all denture wearers. Risk factors for development of denture stomatitis include poor oral/denture hygiene, as well as prolonged wearing of dentures. The denture itself can act as a reservoir for *Candida* through biofilm formation, and can result in a sustained host inflammatory response, therefore regular, thorough cleaning of the denture is essential to prevent the development of denture stomatitis (Jabra-Rizk *et al.*, 2016, Vila *et al.*, 2020).

Candida infection and overgrowth in the oral cavity also has links to cancers. The formation of biofilm in the mouth can allow other microorganisms to anchor and form more complex and polymicrobial biofilms. One such organism commonly found in polymicrobial biofilms along with *C. albicans* is the Gram-positive bacterium, *Streptococcus mutans*, which has a high carcinogenic potential (Talapko *et al.*, 2021). The build-up of biofilm can also create an anaerobic environment that allows obligate anaerobes to grow unchecked, such as Gram-negative bacterial species *Porphyromonas gingivalis* and *Fusobacterium nucleatum*, while both are normal members of the oral

microbiome, both are thought to be risk factors for progression of cancers, specifically oral squamous cell carcinoma, one of the most common cancers worldwide (Talapko *et al.*, 2021, Nobile and Johnson, 2015, Whitmore and Lamont, 2014). Candidal leukoplakia is another presentation of OC and is most seen in middle aged men who smoke. Leukoplakia consists of raised white plaques on the buccal mucosa and tongue that cannot be scraped off easily, an example is seen in Figure 1.2 (E). Candidal leukoplakia increases the risk of oral cancer, with up to 10% of cases developing into oral squamous cell carcinoma (Vila *et al.*, 2020).

Mild cases of OC are usually treated with topical azoles such as miconazole or clotrimazole, or nystatin for 1-2 weeks. However more severe infections could require oral or intravenous fluconazole treatment (Talapko *et al.*, 2021).

1.1.2 Invasive candidiasis

Invasive candidiasis (IC) is a broad term for *Candida* infections of the blood (candidemia), and of deeper organs and tissue. Invasive candidiasis is often seen after surgery, or in patients who are in intensive care units, recent administration of broad-spectrum antibiotics is also a risk factor (Pappas *et al.*, 2018, Kullberg and Arendrup, 2015). It is estimated that IC can affect up to 250,000 people annually, and can cause up to 50,000 deaths (Kullberg and Arendrup, 2015). The most common species of *Candida* found to cause invasive candidiasis infections is *C. albicans* (Pappas *et al.*, 2018). Indwelling medical devices such as catheters can act as a site for biofilm formation and can lead to candidemia as *Candida* enters the bloodstream (Nobile and Johnson, 2015). About 50% of candidemia infections occur in the ICU setting, where patients often require such indwelling devices (Pappas *et al.*, 2018). Candidemia has a mortality rate of up to 40%, regardless of whether patients are receiving antifungal therapy (Kullberg and Arendrup, 2015).

1.1.3 Vulvovaginal candidiasis

Candida spp. are also found as part of the normal microbial flora of the vagina, where they exist in balance with other microorganisms. In the vaginal mycobiome, *Candida* spp. are the most abundant fungi (Rosati *et al.*, 2020). But, as is similar to OC infection, the overgrowth of *Candida* following changes to the microbiome can lead to infection

(Willems *et al.*, 2020, Goncalves *et al.*, 2016). Vulvovaginal candidiasis (VVC) is the most common infection caused by *Candida* species, with an estimated 138 million people affected annually. It is also thought that the numbers of VVC cases are greatly underreported due to the fact that over-the-counter medications are available (Willems *et al.*, 2020). Typical treatment for VVC involves topical or oral antifungal therapy, such as fluconazole, itraconazole or nystatin (Goncalves *et al.*, 2016).

Seventy five percent of women will be affected once in their lifetime, and 8% of women will suffer from recurring vulvovaginal candidiasis (RVVC), meaning they will have more than three cases of VVC per year (Willems *et al.*, 2020). While VVC is not a fatal condition, it increases morbidity for sufferers. Symptoms include vaginal itching, burning and pain. In contrast to OC and IC, VVC can cause infection regardless of immune status, with the immunocompetent and otherwise healthy being affected. It is estimated that by 2030 up to 158 million people could be affected, with a predicted loss in productivity amounting to $14.39 billion (USD) annually (Willems *et al.*, 2020, Rosati *et al.*, 2020, Rosentul *et al.*, 2014).

Similar to OC and IC, the causative species are typically *C. albicans, C. glabrata, C. tropicalis, C. parapsilosis* and *C. krusei*, with *C. albicans* being the cause of around 85% or more of cases. Again, like OC and IC, the percentage of cases being caused by non-*albicans* species has been steadily increasing (Goncalves *et al.*, 2016). It is thought that the increased use of azole drugs has controlled *C. albicans* infections but allowed for the rise in non-*albicans* infections due to the intrinsic resistance to azoles possessed by these species (Goncalves *et al.*, 2016). In cases of RVVC it is thought that non-*albicans* species may be more likely to cause recurrent infections due to their resistance to azole treatment and their ability to recover more quickly after treatment (Rosati *et al.*, 2020).

Risk factors for VVC infection include antibiotic treatment and high oestrogen oral contraceptives, as well as pregnancy, sexual intercourse and uncontrolled diabetes (Willems *et al.*, 2020). Risk factors for RVVC are still unclear, but many believe host genetic factors play a role in predisposition to these recurring infections. For example, polymorphisms in the host *TLR2* gene that decrease the immune response to fungal pathogens are thought to increase susceptibility to RVVC (Rosentul *et al.*, 2014).

1.2 *Candida albicans*

As discussed above, *C. albicans* is the most common cause of human candidiasis infections. The *C. albicans* genome was one of the first eukaryotes to be sequenced, with the diploid sequence being published by Jones *et al.* in 2004 (Jones *et al.*, 2004). The *C. albicans* genome is made up of eight diploid chromosomes, numbered 1-7 and R, the total genome size is 14,851 kb. There are 6,419 ORFs that are larger than 100 amino acids (Odds *et al.*, 2004). There are still many uncharacterised ORFs in the *C. albicans* genome. According to the *Candida* Genome Database (Skrzypek *et al.*, 2017) (on 12/08/2021), 69.27% (4307) of ORFs remain uncharacterised, a further 2.44% (152) are of dubious characterisation, leaving only 28.29% (1759) of ORFs characterised and verified.

Candida albicans is both a successful coloniser and an adaptive opportunistic pathogen, and the fact that there appears to be no environmental reservoir for this yeast indicates that *C. albicans* has adapted specifically to the mammalian host (Noble *et al.*, 2017). It is believed that the phenotypic plasticity that *C. albicans* can display is one of the keys to its success as a commensal in the human host, as well as allowing it to shift to a pathogenic phenotype when conditions are right. Up to 12 different morphotypes of *Candida albicans* cells have been described. Switching between cell types is a response to environmental factors, and allows *C. albicans* to survive in different anatomical environments, to invade host cells and can help evade the host immune response (Gow and Yadav, 2017). This morphogenic plasticity is one of many fitness attributes possessed by *C. albicans*. More in-depth descriptions of these attributes and virulence factors can be found in the sections that follow.

1.3 *Candida dubliniensis*

Previously misidentified as *C. albicans*, a novel species of *Candida* was identified in HIV-patients with OPC in Dublin, Ireland in 1995. The subsequently named *Candida dubliniensis* was distinguished from *C. albicans* due to differences in phenotypic and genotypic characteristics (Sullivan *et al.*, 1995). The whole genome sequence of *C. dubliniensis* was published in 2009. The genome is around 14,600 kb, and the karyotype is more fragmented than that of *C. albicans* (Magee *et al.,* 2008, Jackson *et al.,* 2009).

Estimates suggest that *C. albicans* and *C. dubliniensis* diverged about 20 million years ago, and genomic comparison of the genomes show 96.3% similarity, with almost 80% identity between genes and 98% of genes being syntenic (Moran *et al.*, 2012). While they are closely related genetically, many differences exist between *C. albicans* and its less pathogenic relative *C. dubliniensis*. *Candida dubliniensis* is seen much less frequently in human infection, and it has been shown in infection models that *C. dubliniensis* is the least virulent of the two species (Moran *et al.*, 2012). In the case of OPC, a review of 14 separate studies between 2008 and 2017 found that *C. albicans* was the most common cause of OPC, with frequencies of between 37.2 – 95.2%. *C. dubliniensis* was observed in between 1.48 – 48.9% of cases, and was reported in 9 of the 14 studies reviewed (Patil *et al.*, 2018). *Candida dubliniensis* is also a much less common member of the human oral microbiota (Loreto *et al.*, 2010). *Candida dubliniensis* typically grows as a yeast, and although it can form true hyphae, the responses of *C. dubliniensis* to environmental conditions require stronger stimuli than *C. albicans*. This could be due to differences in the signalling pathways activating transcription necessary for these responses. It has been seen that *C. dubliniensis* forms fewer hyphae *in vivo*, and this could also be a reason it is less virulent. Nutrient depletion is required for a strong hyphal response in *C. dubliniensis*, however, *C. albicans* can form hyphae regardless of nutrient availability (O'Connor *et al.*, 2010). The apparently less effective yeast-hyphal switching system in *C. dubliniensis* may explain why fewer infections are caused by this fungus compared to the number of those caused by *C. albicans*, since the yeast-hyphal switch has been shown to be essential to *C. albicans* virulence, discussed in more detail in Section 1.4 (Noble *et al.*, 2017).

Comparative genomic analysis has revealed that the most significant differences between *C. albicans* and *C. dubliniensis* have been found in gene families related to virulence, with many of these differences being in the number of family members in each species. For example, there are eight genes that are part of the *ALS* family of agglutinin-like sequences in *C. albicans,* while there are only six in *C. dubliniensis* (Jackson *et al.*, 2009). The Als proteins can act as adhesins which are important for host cell attachment. An important note here is that *C. albicans* is the only *Candida* species to possess the *ALS3* gene, which is expressed specifically on the surface of hyphae. The Als3 protein can bind to host cell receptors, N-cadherin and E-cadherin and induce endocytosis, it can also bind ferritin in the host and use protein as an iron source (Liu and Filler, 2011).

The *IFA* family of leucine rich repeats are also expanded in *C. albicans* compared to *C. dubliniensis.* In *C. albicans* there are 31 *IFA* genes, however, six are thought to be non-functional. In *C. dubliniensis* there are 21 *IFA* genes, and in this case 14 are thought to be non-functional. There is one single *IFA* gene in *C. tropicalis* (McManus and Coleman, 2014, Jackson *et al.*, 2009).

The *SAP* family of secreted aspartyl proteinases are also expanded in *C. albicans* compared to *C. dubliniensis* (and other *Candida* species) (Jackson *et al.*, 2009). *Candida albicans* has 10 different *SAP* genes, and *SAP4, 5* and 6 have been shown to be necessary for systemic infection in mouse models and are only expressed by hyphae. *SAP5* and *SAP6* are missing from the *C. dubliniensis* family (Moran *et al.*, 2012).

The Telomere-associated ORF (*TLO*) family is a family of Med2 homologues that incorporate into the Mediator complex. This family is highly expanded in *C. albicans* compared to other *Candida* species. *Candida dubliniensis* has two *TLO* genes, which are both divergent from those in *C. albicans*. Other *Candida* species only possess one single *TLO* gene (Jackson *et al.*, 2009). The *TLO* gene family is discussed in Section 1.5.

1.4 Phenotypic plasticity of *Candida albicans* and *Candida dubliniensis*

Previously referred to as dimorphic, *C. albicans* is more accurately referred to as a polymorphic fungus. Like other dimorphic fungi it can switch between a yeast and filamentous hyphal morphology, Figure 1.3 (A) and (C) respectively. However, several other cell types have been identified. Many of these novel cell types are yeast-like and include the opaque cell type which is mating competent and found to be involved in colonisation of certain organs, like the heart, spleen, and skin (Noble *et al.*, 2017, Takagi *et al.*, 2019), and gastrointestinally induced transition cells which are elongated yeast-like cells that have been found to outcompete other cell types in gastrointestinal colonisation (Noble *et al.*, 2017).

Candida albicans and *C. dubliniensis* are the only *Candida* species that can form chlamydospores, which are thick-walled suspensor cells that grow on the ends of filamentous protrusions in nutrient limiting conditions (Nobile *et al.*, 2003). The role of chlamydospore formation in the *Candida* life cycle is still unclear, but it is believed that the conservation of chlamydospore formation means it is biologically important

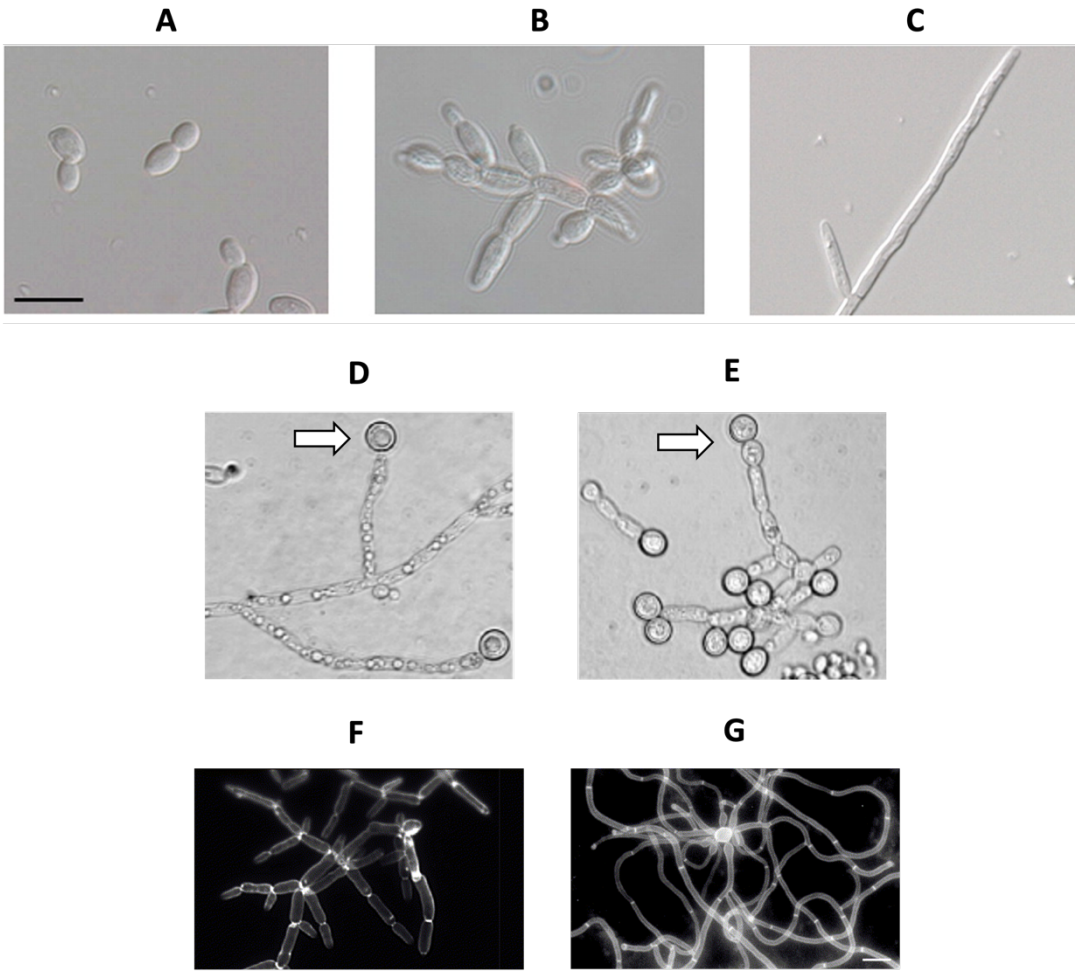


Figure 1.3 Cellular morphologies of *Candida* species

(A) Yeast cell morphology (blastospore) of *C. albicans* (Thompson *et al.*, 2011), scale bar represents 10 μm with scale being equal across A, B and C. (B) Pseudohyphal cell morphology of *C. albicans* (Thompson *et al.*, 2011). (C) True hyphal morphology of *C. albicans* (Thompson *et al.*, 2011). (D) Chlamydospore, indicated by white arrow, in *C. albicans* (Staib and Morschhaeuser, 2007). (E) Chlamydospore, indicated by white arrow, in *C. dubliniensis* (Staib and Morschhaeuser, 2007). (F) Pseudohyphal morphology after staining with Calcofluor White (Sudbery *et al.*, 2004). (G) True hyphae stained with Calcofluor White (Sudbery *et al.*, 2004). Cell walls and septa are stained with dye, highlighting morphological differences between true hyphae and pseudohyphae. Scale bar represents 10 μm, with both F and G being to the same scale.

(Bottcher *et al.*, 2016). There are some differences in the chlamydospore morphologies of *C. albicans* and *C. dubliniensis*, *C. albicans* forms a single suspensor chlamydospore at the end of a filament, see Figure 1.3 (D), while *C. dubliniensis* can produce multiple chlamydospores at the end of a single filament, Figure 1.3 (E), (Campanha *et al.*, 2005). Chlamydospore formation can be used as a differential test to determine if an isolate is *C. albicans* or *C. dubliniensis*, where only *C. dubliniensis* forms chlamydospores on Staib agar (Staib and Morschhaeuser, 2007). It is believed that many of the genes involved in the hyphal development pathway are involved in the chlamydospore formation pathway, with Nrg1 appearing to have the most important role. Nrg1 is a hyphal repressor and seems to also play a role in repressing chlamydospore formation in *C. albicans*. Differences in Nrg1 activity between *C. albicans* and *C. dubliniensis* is thought to be responsible for the difference in chlamydospore formation in these species (Bottcher *et al.*, 2016). Arguably the most important phenotypic switch in *C. albicans* morphology is the yeast-hyphal switch. *Candida albicans* and *C. dubliniensis* are the only two *Candida* spp. that can form true hyphae (Sudbery *et al.*, 2004). This switch is induced in response to a wide variety of environmental cues, such as changes in pH, temperature, or availability of nutrients. The ability to respond in such a way to environmental conditions is key to the ability of *C. albicans* to survive in various, distinct locations in the body, and it is thought that this feature is an essential part of its pathogenesis. Hyphal cells are able to damage the epithelium by invasion and by growing between cell junctions, they can puncture and damage endothelial cells, and during a host immune response, hyphal cells can evade neutrophils and macrophages by lysis (Thompson *et al.*, 2011).

In mucosal infections, the hyphal form is required for invasion of the epithelial cells. In bloodstream infections, it is the yeast form that is prevalent (Sudbery *et al.*, 2004), but it has been shown that mutant strains which do not have the ability to switch between yeast and hyphal forms cannot successfully cause a bloodstream infection in a murine model (Noble *et al.*, 2017). Therefore, it is not only the hyphal or the yeast morphology that is important in causing infection, but the morphological plasticity itself that is required for pathogenesis.

Morphologically, yeasts are oval and unicellular, they divide by budding. Hyphal cells grow as elongated filaments which are uniform in width with parallel walls and have septa dividing different cells within the hyphae, however, these septa have pores that allow cells within hyphae to communicate (Thompson *et al.*, 2011). Figure 1.3 depicts

hyphal cell growth and highlights these septa within the filaments. When hyphal growth is induced, the initial hyphal filament that extends from the yeast cell, before formation of the first septum, is termed a germ tube. The induction of germ tube formation is used as a diagnostic test in the clinical setting to confirm *C. albicans* infection, with *C. albicans* and *C. dubliniensis* being the only two *Candida* spp. forming germ tubes in response to serum (Sudbery *et al.*, 2004). Hyphal cells also contain a specialised organelle called a Spitzenkörper, which is present at the apical tips of hyphae that directs growth (Thompson *et al.*, 2011).

Another filamentous morphology of *C. albicans* exists, termed the pseudohyphal form. This form resembles a mid-way point between yeast and hyphal cells. Pseudohyphae occur when a daughter cell elongates and begins to separate from the mother cell, but after septum formation, the cells do not detach. This results in highly branched elongated cells, much more branched than true hyphae, which are not uniform in shape and do not have parallel sides (Sudbery *et al.*, 2004). Pseudohyphal cells are similar to hyphal cells in that they are elongated, sometimes to a dramatic degree, but they are not truly hyphal as they lack the pored septa separating the individual cells (Thompson *et al.*, 2011). A comparison of these cell types can be found in Figure 1.3.

The transition between yeast and hyphal cell morphology is controlled by a complex network of signal transduction pathways, some of which overlap, but in a general sense, hyphal growth is induced by upregulation of hypha-associated genes (HAGs) and yeast-like growth is induced the expression yeast specific genes. It is not the case that activation of all HAGs is required to turn on hyphal growth, therefore these genes are termed hypha-associated, and are not necessarily hypha-specific (Basso *et al.*, 2019).

There are many different environmental factors that can stimulate hyphal formation. These include pH higher than 7, temperature of 37 °C, the presence of serum, oxidative stress and quorum sensing (Mayer *et al.*, 2013). As mentioned previously, the formation of germ tubes by *C. albicans* in serum at 37 °C is used as a diagnostic test in the clinical setting. Some of the genes and pathways involved in hyphal initiation are described below.

The cAMP-PKA pathway can be induced by temperature and pH signals, among others. This pathway involves the synthesis of cyclic AMP which activates a protein kinase A complex (PKA) that phosphorylates the transcription factor enhanced filamentous

growth protein 1 (Efg1). Efg1 is a key transcription factor in upregulating HAG expression and is implicated in several pathways across the network (Noble *et al.*, 2017, Sudbery, 2011). An Efg1 null mutant strain was shown to be locked into a yeast morphology, and hyphal growth could not be induced (Banerjee *et al.*, 2008). Efg1 is responsible for HAG induction in the RIM101 signalling pathway also, which is a proteolytic signalling cascade induced by pH (Noble *et al.*, 2017).

Mitogen-activated protein kinase (MAPK) signalling pathways can induce HAG expression and can be activated by several environmental signals. The Cek1 MAPK pathway is activated by signals such as call wall damage, osmotic damage or nitrogen starvation. This pathway activates the expression of HAGs via the Cph1 transcription factor (Noble *et al.*, 2017). Repression of HAGs is facilitated by Tup1, a general transcriptional repressor that can be targeted to HAGs by DNA-binding protein Nrg1 (Sudbery, 2011).

Ume6 is regulator of hyphal extension. A *ΔΔume6* mutant strain was shown to be able to form germ tubes, but to be unable to produce true hyphae. It is understood that Ume6 also works to repress Nrg1 and prevent repression of HAGs to maintain hyphal growth (Banerjee *et al.*, 2008). The Brg1 transcription factor is also involved in the repression of Nrg1 mediated HAG repression, resulting in hyphal growth. Brg1 recruits Hda1, a histone deacetylase, to HAG promoters, which modifies chromosome structure and prevents binding of Nrg1 and subsequent repression (Basso *et al.*, 2019).

1.4.1 Phenotypic testing in Candida *species*

Phenotypic testing is important for characterising *Candida* strains, and to determine the effects of gene deletions and mutations. There are a wide variety of tests that investigate many aspects of *Candida* biology, and here the significance of the tests used in this work will be briefly described.

Growth rate analysis is performed to investigate the metabolic fitness of *Candida* strains, and their ability to utilise carbon sources. Strains with high growth rates are typically deemed more fit, and the increased rate of proliferation may play a role in pathogenesis (Dunker et al., 2021). A typical carbon source used for these types of analyses is glucose, however a wide variety of carbon sources can be used to investigate different metabolic pathways. In this work, galactose is used as an alternative carbon

source for growth rate analysis as previous work highlighted that growth of a *Δtlo1/Δtlo2* strain of *C. dubliniensis* was significantly slower than WT, indicating a role for Tlos in the galactose metabolism pathway (Haran et al., 2014). Various temperatures can also be used in growth rate assays, 37 °C being represents the temperature encountered when inhabiting the human host, 30 °C is also frequently used for growth rate analysis. To examine the ability of strains to grow in high temperatures, assays can be performed at 42 °C, and previous studies have shown that *C. dubliniensis* does not grow well at this temperature, while *C. albicans* can grow successfully (Sullivan and Coleman, 1995). The use of the respiration inhibitor Antimycin A can be used to examine the ability of strains to enlist alternative pathways to grow, such as the alternative oxidase pathway (Chabrier-Rosello et al., 2010).

The ability of *C. albicans* and *C. dubliniensis* to form hyphae is crucial for pathogenesis and has been discussed above. Phenotypic assays in vitro examining the ability of mutant strains to form hyphae can help uncover if deleted genes are involved in the yeast to hyphal morphogenesis pathway. Similarly tests that induce chlamydospore formation can also highlight defects in these pathways in mutant strains. A *Δtlo1/Δtlo2* strain of *C. dubliniensis* was unable to form hyphae and chlamydospores in phenotypic testing and indicated a role for Tlos in these pathways (Haran et al., 2014).

Biofilm formation is an important phenotype for *Candida* species, as it allows persistence in the environment, as well as being a reservoir for human infection (e.g. biofilm formation on indwelling medical devices progressing to infection). *Candida* biofilms form when yeast cells adhere to a surface and proliferate into a base layer, following this, hyphae and pseudohyphae grow on top of this layer and form a complex biofilm surrounded by an extracellular matrix. Yeast cells can then disperse and form new biofilms at other sites (Nobile et al., 2012). Phenotypic assays measuring the extent of biofilm formation in *Candida* mutants can identify strains that may have defective biofilm formation pathways and may highlight genes that are important in the biofilm formation pathways. Deletion of *TLO* genes from *C. dubliniensis* was found to affect the ability of this strain to form biofilm (Haran et al., 2014).

The fungal cell wall is a very important organelle as it is the interface between the cell and the external environment. Cell wall biosynthesis is complex, it involves many genes and requires a lot of energy (Ram and Klis, 2006). Defects in the *Candida* cell wall or in cell wall biosynthesis can result in major defects in the cell, such as reduced growth rate, increased sensitivity to external stress and attenuated virulence (Garcia-Rubio *et al.,*

2020). Assays that target the cell wall and test the sensitivity to cell wall perturbing compounds are important tools for identifying strains which have defects in cell wall synthesis pathways. Such compounds include Calcofluor White and Congo Red, which target β-linked glucans in the cell wall (Ram and Klis, 2006). *Δtlo1/Δtlo2* strain of *C. dubliniensis* was found to be sensitive to these compounds (Haran et al., 2014).

Host immune cells produce reactive oxygen species (ROS) to kill invading pathogens, and the ability of *Candida* cells to resist this stress contributes to the progression of infection. In vitro H_2O_2 is used to test the susceptibility of strains to oxidative stress. However, this chemical can be unstable over long periods of time and so *tert*-butyl hydroperoxide (tBOOH) can be used as a more stable alternative (Kaloriti *et al.,* 2012). *Δtlo1/Δtlo2* strain of *C. dubliniensis* was found to be sensitive to oxidative stress (Haran et al., 2014).

The ability of *Candida* mutant strains to resist and tolerate antifungal drugs is routinely examined as part of the phenotypic testing of strains. Microbroth dilution assays, spot plate assays and the use of Etests are common methods used to determine the minimum inhibitory concentration of drug for Candida strains. The most common class of antifungal drug used is the azole class (e.g. fluconazole). Echinocandins (e.g. amphotericin B) and polyenes (e.g. anidulafungin) are also commonly used antifungals and are also used for drug sensitivity testing (Liu et al. 2005, Vasicek et al. 2014). Drug sensitivity testing was not performed as part of this thesis, as this work was being performed by Dr James O'Connor-Moneley as part of a postdoctoral research project.

In terms of infection models, the murine model is the gold standard for investigating the virulence of pathogens, however the use of this model requires ethical approval and specialised training, as well as specialised laboratory facilities to house and care for the animals. Many labs do not have these facilities, and so alternative infection models are required. Human macrophage cells can also be used to study the interaction of *Candida* and host cells (Lorenz et al., 2004), however, this again requires specialised cell culture facilities which not all laboratories have access to. One alternative, the greater wax moth larvae (*Galleria mellonella*) model, is used in this work. Lack of requirement for ethical approval, the relatively low cost and ease of purchase as well as the simple nature of inoculation and short assay run time (around three days) make this model very accessible for all laboratories (Pereira et al., 2015). This model is a simple method for assessing the virulence of *Candida* strains over a relatively short period.

1.5 The *TLO* gene family

The *TLO* gene family is highly expanded in *C. albicans*, and one hypothesis is that expansion of this gene family could be a contributing factor to the phenotypic plasticity and virulence of this species. The *TLO* genes are Med2 homologues and Tlo proteins can incorporate into the Mediator complex, see Section 1.6. They are similar to the *MED2* gene in *S. cerevisiae*. The expansion of the *TLO* gene family in *C. albicans* is unique, other *Candida* spp. only have one *TLO* gene, and *C. dubliniensis* has two, nowhere close to matching the 10-15 *TLO* gene family members that *C. albicans* can have, Figure 1.4. The SC5314 lab strain has 14 *TLO* genes, and details about these genes can be found in Table 1.1, including their genomic location, the sizes of the genes and the sizes of the proteins they encode. There is also a pseudogene, *TLOψ4,* which lacks the Med2 domain. Details about the *C. dubliniensis TLO* genes can also be found in this table.

The expansion of the *TLO* gene family in *C. albicans* is the largest expansion of gene copy number between *C. albicans* and other members of the CTG clade (Zhang *et al.*, 2012, Hirakawa *et al.*, 2015). The *C. albicans TLO* genes are less related to *C. dubliniensis* or *C. glabrata* than these two are to each other, suggesting that the *TLO* genes in *C. albicans* has undergone more modification than those of any other *Candida* species. *Candida albicans* has derived novel copies of the *TLO* genes, presumably though duplication at its telomeric regions, possibly though telomeric exchange, or retrotransposon activity (the LTRs that are used as a basis to divide the *TLO*s into clades, which is discussed below, is a product of retrotransposon activity) (Jackson *et al.*, 2009, Sullivan *et al.*, 2015). It appears that translocation of the *MED2* gene into the *Candida* ancestor occurred very early in the evolution of the genus. The position of the *CaTLOβ2* gene appears to be the original site of this translocation, as there are *TLO* homologues found in this same position in all related *Candida* spp. (Sullivan *et al.*, 2015). Most subtelomeric *TLO*s, with the exception of *TLOβ2* and *TLOγ16* and the non subtelomeric *TLOα34*, are flanked on their 3' end by a *TLO* Recombination Element (TRE) which extends into the Bermuda Triangle Sequence (BTS); these are highly repetitive sequences that are thought to play roles in subtelomeric recombination (Dunn and Anderson, 2019). It is hypothesised that the expansion of this gene family may contribute to the increased virulence of *C. albicans*. It is proposed that when different Tlo proteins associate with Mediator in different complexes they may affect different responses in the organism, such as yeast to hyphal switch, or expressing proteins that aid in pathogenesis (Zhang *et al.*, 2012). See Chapter 5 for more on *TLO* clades and the expression of these genes.

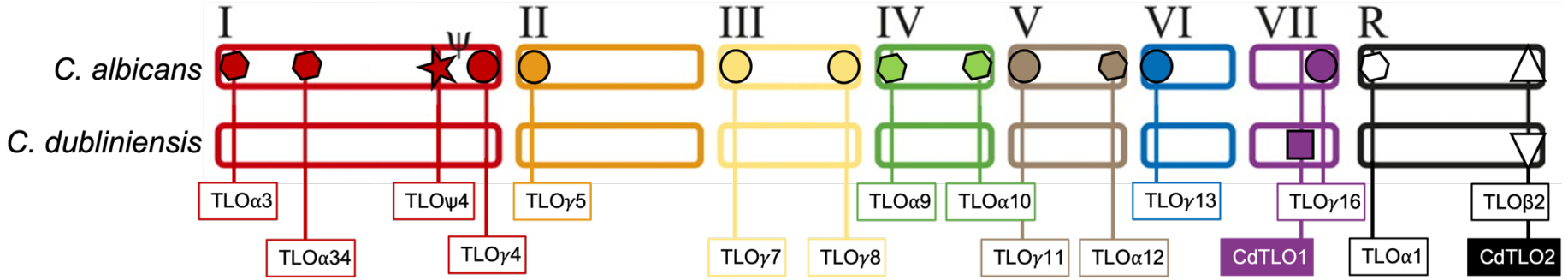


Figure 1.4 *TLO* distribution across chromosomes in *C. albicans* and *C. dubliniensis*

Here, the *TLO* genes of C. albicans and *C. dubliniensis* are mapped onto the chromosomes, *C. albicans* on top and *C. dubliniensis* on the bottom. Karyotyping is based on *C. albicans*, it is not correct for *C. dubliniensis* and is for visualisation purposes only. Chromosome numbers for *C. albicans* are indicated by Roman numerals for chromosomes 1-7 and R for the R chromosome. Each chromosome is represented by a different colour also. The *TLO* gene locations on each chromosome are indicated by a symbol, and a text box linked to each symbol describes the gene. Symbols represent the clade of *TLO* gene; hexagon = α, triangle = β, circle = *γ* and the pseudogene TLOψ4 is marked with a star. There are no clades dividing *C. dubliniensis TLO* genes and so two random symbols mark the positions of these genes. This figure was adapted from (Jackson et al., 2009).

Candida albicans

Protein	Gene	Clade	Systematic Name	Assembly 21 Identifier	Chromosome	Coordinates	Size (bp)	MW (kDa)	Subtelomeric
Tlo1	*TLO1*	α	CR_00020W_A	orf19.7544	R	9111-9863	753	28.844	✓
Tlo2	*TLO2*	β	CR_10860C_A	orf19.7680	R	2285377-2286198	822	31.272	✓
Tlo3	*TLO3*	α	C1_00040W_A	orf19.6112	1	10718-11485	768	29.263	✓
Tlo34	*TLO34*	α	C1_06180W_A	orf19.2661	1	1292811-1293806	996	37.547	×
Tlo4	*TLO4*	γ	NOT IN CGD	NOT IN CGD	1	3187464-3187887	432	16.08	✓
Tlo5	*TLO5*	γ	C2_00010W_A	orf19.1925	2	4248-4778	531	20.836	✓
Tlo7	*TLO7*	γ	C3_00060W_A	orf19.5467	3	13756-14265	510	19.57	✓
Tlo8	*TLO8*	γ	C3_07970C_A	orf19.6191	3	1787576-1788085	510	19.867	✓
Tlo9	*TLO9*	α	C4_00010W_A	orf19.362	4	983-1660	678	25.578	✓
Tlo10	*TLO10*	α	C4_07250C_A	orf19.3074	4	1597159-1597812	654	24.72	✓
Tlo11	*TLO11*	γ	C5_00010W_A	orf19.5700	5	1918-2427	510	19.993	✓
Tlo12	*TLO12*	α	C5_05500C_A	orf19.4054	5	1182110-1182868	759	28.742	✓
Tlo13	*TLO13*	γ	C6_00030W_A	orf19.6337	6	5545-6069	525	20.49	✓
Tlo16	*TLO16*	γ	C7_04360C_A	orf19.7127	7	942825-943352	528	20.467	✓

Candida dubliniensis

Protein	Gene	Systematic Name	Chromosome	Coordinates	Size (bp)	MW (kDa)	Subtelomeric
Tlo1	*TLO1*	Cd36_72860	7	734,111-735,073	963	36.114	×
Tlo2	*TLO2*	Cd36_35580	R	2,254,730-2,253,663	1068	39.773	✓

Table 1.1 Genetic information on Med2 homologues in *Candida albicans* and *C. dubliniensis*

1.6 The Mediator Complex

Mediator is a multiprotein complex that is found in all eukaryotes, serving as a co-activator of transcription. It interacts with RNA Polymerase II, transcription factors bound at regulatory elements and general transcription factors. It is unable to bind directly to specific DNA sequences, however, it affects transcription through physical interaction with RNA Polymerase II, which can be seen in Figure 1.5 (A), (Casamassimi and Napoli, 2007). In *Saccharomyces cerevisiae*, the Mediator complex is required for the transcriptional regulation of almost all RNA Polymerase II dependent genes, both protein coding and non-coding (Tebbji *et al.*, 2014). In fungi, Mediator is made up of 25 subunits, and these are organised into four modules in the structure of the polyprotein – the head which interacts with RNA Polymerase II, the middle, which is involved in regulation, the tail module that has a direct role in transcriptional regulation, and the Cdk module that is variably associated and can both negatively and positively regulate transcription (Sullivan *et al.*, 2015). The composition of the *C. albicans* Mediator can be seen in Figure 1.5 (B). Mediator is a relatively large complex, and this allows it to interact with multiple transcription regulators, simply due to its size. The complex can also adopt distinct conformations at different promoters or adopt activator dependant conformations (Tebbji *et al.*, 2014). The ability of this polyprotein to dramatically change its shape and function likely lends to its role in the regulation of almost all RNA Polymerase II dependent genes (in *S. cerevisiae*). Mediator has been shown to primarily interact with the upstream regulatory sequences (UAS) at yeast promoters via the tail module, which is thought to interact with specific transcription factors at the UAS (Moran *et al.*, 2019).

The C-terminal domain (CTD) of the RNA Polymerase II synchronises transcription and transcription-associated processes by acting as a signalling platform, which, through a cycle of phosphorylation and dephosphorylation, can recruit transcription-associated factors (Bartkowiak and Greenleaf, 2011). Mediator forms a holoenzyme with RNA Polymerase II through direct interaction between the Mediator and the CTD of RNA Polymerase II. Once transcriptional elongation begins, Mediator dissociates from RNA Polymerase II, and the polymerase carries on without it. Mediator can only associate with the CTD in the unphosphorylated form, and cannot bind to a phosphorylated CTD. It is believed that the Cdk subunit plays a role in phosphorylation on the CTD domain of RNA Polymerase II, and thus regulates transcription (Casamassimi and Napoli, 2007). The role of Mediator in *C. albicans* is less well documented than that in *S. cerevisiae*.

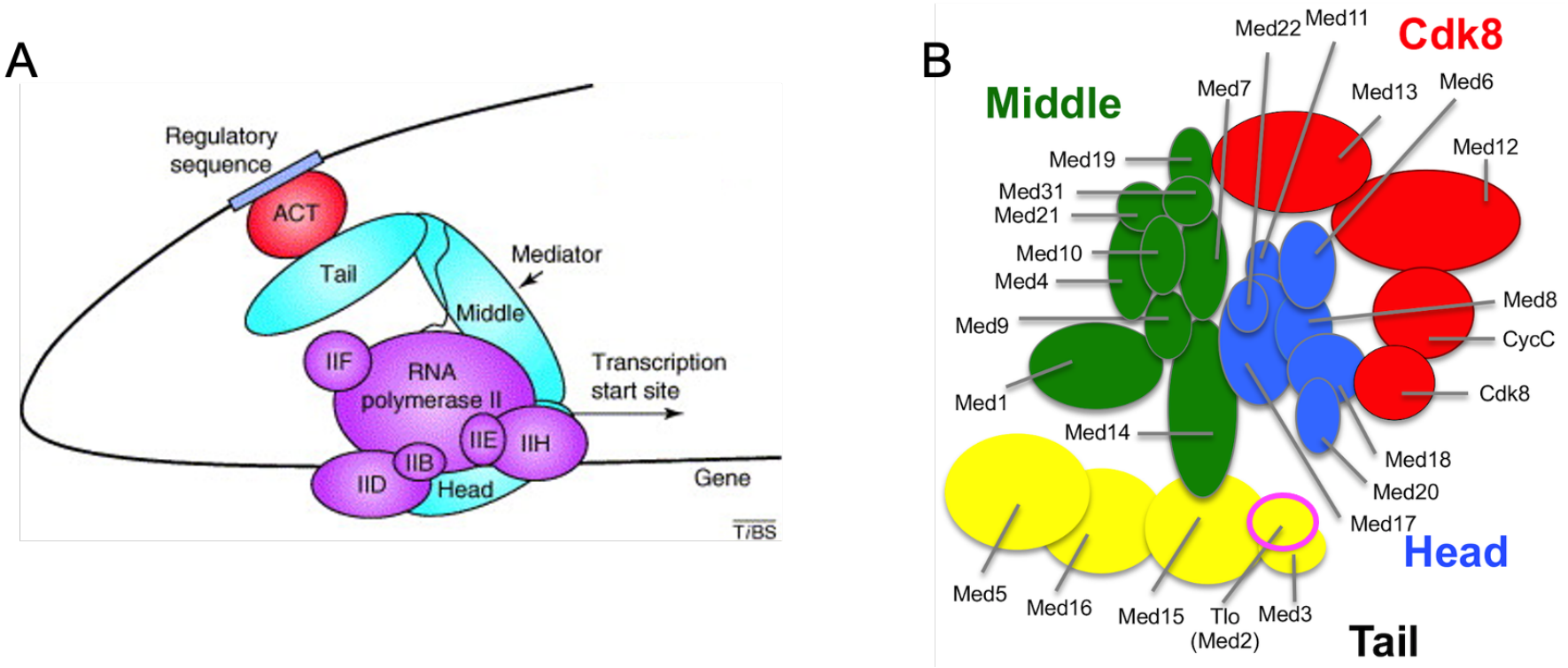


Figure 1.5 The Mediator Complex

(A) The Mediator complex interacts with RNA Polymerase II to drive transcription. The head of the complex interacts with the polymerase, while the tail plays a role in directing the complex to upstream activation sequences, where it interacts with transcription factors, driving expression of specific genes. This diagram is from (Bjorklund and Gustafsson, 2005). (B) The *C. albicans* Mediator complex is made up of 25 subunits, these are organised into head, middle, tail and Cdk8 modules. The Tlo proteins are Med2 homologues and integrate into the Mediator at the Tail module in a Med3 dependent manner, the Tlo protein can be seen in the figure highlighted in a pink circle. This figure is adapted from (Sullivan *et al.*, 2015).

Some known functions of Mediator subunits in *C. albicans* include roles in expression of virulence traits. Med31 (middle) and Med20 (head) have been implicated in biofilm production and also have a role in the yeast-hyphal morphology switch. Med31 has been proven to be required for expression of *ALS1* and *ALS3* genes, which encode cell surface proteins essential in biofilm formation. Mutants with Med31 deleted are unable to produce biofilms. Med31 mutants are also unable to undergo proper cytokinesis, leading to defective cell wall remodelling during cell separation (Uwamahoro *et al.*, 2012). In terms of essential subunits, only deletions of subunits in the head region or the middle region have been shown to be non-viable. Deletion of *MED4, 10, 14* or *21*, which incorporate into the middle, or deletion of *MED8, 11* or 17, which incorporate into the head, each proved non-viable (Sullivan *et al.*, 2015). Mediator subunits each have unique roles, and disruption of any of these subunits will understandably have dramatic effects on the transcriptional pathways in the cell. In fungi, half of all Mediator subunits are encoded by essential genes, and these have about 20-30% amino acid identity with their counterpart subunits in mammals. However, the Med2 and Med3 subunits are highly divergent in Ascomycetes, with no counterparts in mammals (Liu *et al.*, 2016). Tlo proteins, which are Med2 homologues, and Med3, along with Med15, comprise the tail module of the Mediator complex. It was found that the Tlo proteins did not incorporate into the Mediator complex of a *ΔΔmed3* mutant strain of *C. albicans*, indicating that anchoring of the Tlo proteins in Mediator is dependent on Med3 (Zhang *et al.*, 2012). The position that Tlo proteins occupy in the Mediator complex, and their interaction with Med3 can be seen in Figure 1.5 (B), where the Tlo protein is highlighted. Tlo proteins in *Candida* species are thought to have significant roles in transcriptional regulation of virulence factors, particularly in *C. albicans* and, to a lesser extent, in *C. dubliniensis* (Haran *et al.*, 2014).

1.6.1 Mediator free Tlo

In *C. albicans,* while the Tlo proteins do incorporate into the Mediator complex, there is also a large pool of unincorporated, 'free' Tlo. In fact, the majority of Tlo protein in *C. albicans* is in this free state, with up to 10 times more Tlo existing outside of Mediator than within it (Liu *et al.*, 2016). In *C. dubliniensis*, CdTlo1 exists only as a part of the Mediator complex, and cannot be detected in an unassociated form, even when artificially overexpressed. While under normal conditions *CdTLO2* is expressed at around 50 times less than *CdTLO1* mRNA and exists at around 30 times less on the

protein level. It is possible to artificially generate a pool of free CdTlo2 protein in the *C. dubliniensis* cell. This overexpression was seen to enable *C. dubliniensis* filamentation on a scale similar to that of *C. albicans* (Liu *et al.*, 2016). Generation of this free pool of Tlo protein in *C. dubliniensis* demonstrates the possible consequences of the free Tlo pool in *C. albicans* and shows that free Tlo could be enacting functions outside of association with Mediator.

C. albicans α and β clade Tlos have been shown to contain a transcriptional activation domain (TAD) which can function independently to Mediator (Liu *et al.*, 2016). These TADs have also been observed in *C. dubliniensis* Tlos and the Med2 protein in *S. cerevisiae*. It is thought that through these TADs the free Tlo proteins, acting independently of Mediator, can compete for DNA binding sites on co-activators and co-repressors, enacting transcriptional changes on their own accord (Liu *et al.*, 2016). The large pool of free Tlo present in *C. albicans* may also be 'squelching' the response of the Mediator complex, where Tlo proteins sequester DNA binding sites on co-activators that Mediator is targeting via the tail module, and so blocking transcriptional activation (Zhang *et al.*, 2012).

1.7 Aims of This Work

The work carried out within this thesis aimed to characterise the role of the *TLO* genes of *C. albicans* with regards to function within the cell, and the role of their expansion in *C. albicans* virulence compared to other *Candida* species. Dose dependent effects and clade specific effects were investigated via expression of *C. albicans TLOs* in a *C. dubliniensis* wild type strain to determine if expansion of the *TLO* gene family in *C. dubliniensis* can increase virulence of this species. A *TLO* null mutant strain of *C. albicans* was generated via CRISPR-Cas9 mutagenesis. This strain was used as a background to test the clade specific and effects of the expression of representative *TLO* genes without interference from other family members. This background was also used to determine the effects of overexpressing single *TLOs* in a depleted background. Next generation sequencing analysis was performed on these strains. RNA-sequencing was used to examine the transcriptional profiles of the null mutant and reintegrated strains, under various conditions. ChIP-sequencing was used to determining potential interaction sites for Tlo proteins on the DNA. The results of these experiments are discussed in the subsequent chapters, and a discussion on the significance of these data follows.

Chapter 2

General Materials and Methods

2.1 General Materials and Methods

2.1.1 Buffers and solutions

All of the buffers and solutions used in this work were of molecular biology grade. Ultrapure Milli-Q Biocel purified water (Millipore, Carrigtohill, Cork, Ireland) was used to make up all solutions and buffers. Molecular biology grade water (Sigma-Aldrich, Arklow, Co. Wicklow, Ireland) was used to dilute DNA and oligonucleotides, and was also used in all molecular reactions (e.g. polymerase chain reactions, restriction endonuclease digests, etc.). Phosphate Buffered Saline (PBS) 1X stock was made from Oxoid tablets (Oxoid, Basingstoke, Hampshire, UK). Tris-Borate EDTA 5X concentrate sachets (Sigma-Aldrich) were used to make up 0.5X TBE buffer used as a buffer for running gel electrophoresis and also for dissolving agarose (Sigma-Aldrich).

2.1.2 Chemicals, enzymes, antibiotics and antifungals

Chemicals were acquired from Sigma-Aldrich (Arklow, Co. Wicklow, Ireland), Roche (Roche Products [Ireland] Limited, Citywest, Dublin 24, Ireland), Promega (Promega Corporation, Madison, Wisconsin, USA), Bioline (London, United Kingdom) and Life Technologies (Dun Laoghaire, Dublin, Ireland). Enzymes were acquired from Promega and New England Biolabs (Ipswich, Massachusetts, USA). All were stored and used according to the manufacturers' instructions. Antibiotics and antifungals were acquired from Sigma-Aldrich, with the exception of nourseothricin (Werner Bioagents [Jena, Germany]).

2.1.3 Oligonucleotides

Unless otherwise stated, all oligonucleotides were purchased from Sigma-Aldrich. Sequences of all oligonucleotides used in this study can be found in Table 2.1. Before use, 100 μM stocks were diluted to 10 μM in Molecular Biology grade water unless otherwise stated.

Name	Sequence
Wü::CaTLOs	
M13	GGAAACAGCTATGACCATG
TAGR	GGGAAATCAACACTTCCAGTG
RT PCR	
ACT1 F	AGCTCCAGAAGCTTTGTTCAGACCAG
ACT1 R	TGCATACGTTCAGCAATACCTGGG
CaTLO1 qPCR F	ACTAGCCCCAACAACGAACT
CaTLO1 qPCR R	CATAACGCCGAGACACCACT
CaTLO2 F qPCR	TCAACGACATGCAGAACGAC
CaTLO2 qPCR R	TCATGTCCAAGTCGCTGTCT
CaTLO11 qPCR F	ATAACCCAACTGCTCAACGG
CaTLO11 qPCR R	CACTTCTTGGCTTCCTCTGC
CaTLO12 qPCR F	AGCAGAAGAAGCAGCAAAGA
CaTLO12 qPCR R	AAGTTGTTGTCTGTGGTGGC
CaCRISPR	
AHO1096	GACGGCACGGCCACGCGTTTAAACCGCC
AHO1098	CAAATTAAAAATAGTTTACGCAAG
AHO1097	CCCGCCAGGCGCTGGGGTTTAAACACCG
MED2 gRNA	CGTAAACTATTTTTAATTTGTTTTTGGATTAGTTCGTTGTGTTTTAGAGCTAGAAATAGC
AHO1237	AGGTGATGCTGAAGCTATTGAAG
AHO1238	TGTATTTTGTTTTAAAATTTTAGTGACTGTTTC
MED2 repair top	TACATAACTCACTCGACGAGATATTGAAATCATCAGGATACATAAATGATGCAGAGTGGTGTCT
MED2 repair bottom	CCATTGTCGAGCAATGTCATGTTCAACCCGTCTAGTATCAGCTCGTCGTTTTCAAGTATAGACACCACTCTGCATCATT
ADE2 gRNA	CGTAAACTATTTTTAATTTGTCTATAGTACAGATGCCAAGGTTTTAGAGCTAGAAATAGC
ADE2 repair top	TAATGGATAGCAAAACTGTTGGTATTTTAGGAGGTTAATGATTAGGTCGTATGATTGTTGAAGCAG
ADE2 repair bottom	CGGTCTTGATATTCAATCTATGTGCTGCTTCAACAATCATACGACCTAAT
TLO checks	
Pan TLO	GACCAGATATTGAAATCATC
TLO1	ACAGGTGGATCAAGTTCTGG
TLO2	GGTCGGTCTTGAATGCTTCG
TLO3	TCCAACTCGTTCTTTAACCTC
TLO4	GCAATCCCATCTGCTATGAC
TLO5	CTCTGGCGATGTATGACGAC
TLO7	CTGTAAGCTGGTTCGAATCT
TLO8	TTCTGGACGGTAGTTGACGG
TLO9	GTCTCTTTCGTGTAAGGTACC

TLO10	AGAAGCCTATGCCTATGCGT
TLO11	CAAGTCTTTGGAGGTGACGG
TLO12	AGTGAGTTCTTCACCTTGG
TLO13	ATCCATCTACCACTGCTGCC
TLO16	CACCATAAGATGCTGGAAGG
TLO34 F	ATACCTTTCTTTTCTCTTCT
TLO34 R	CTATCTGTATATGTCTTGTAG
Cloning	
TLO34 3Prime (*Sac*II) F	ATGCCCGCGGAGGGCTCGTTCTACAAGAC
TLO34 3Prime (*Sac*I) R	ATGCGAGCTCTTGTAGAGTCGTGGCATTCG
TLO (*Sal*I) F	ATGCGTCGACACCATGCCAGAAAACCTCCA
TLO tag (*Bgl*II) R	ATGCAGATCTACCTAAGCGTAATCTGGAAC
Reintroduction Checks	
TLO34 locus F	TCGATTTCTACTTGATGCACGA
TLO1 R	ACTTTTGTGGGACGGTCTCT
TLO2 R	GTCGTTCTGCATGTCGTTGA
TLO11 R	CACTTCTTGGCTTCCTCTGC
ADH1 locus F	CATGTGCACGGACAAGCTTA
carTA R	AATACTTTATAAAAGCTAAGCTA

Table 2.1 Oligonucleotide primers used in this work

Table of oligonucleotide primer sequences used in this work. Red bases in gRNA sequences indicate target sequences of that guide. Underlined bases indicate restriction enzyme recognition sites.

2.2 Microbial Strains

2.2.1 Strains and isolates

Strains used in this study, unless otherwise stated, were all obtained from the Dublin Dental University Hospital (DDUH) strain collection at the DDUH, University of Dublin, Trinity College, Lincoln Place, Dublin 2. All strains used in this study, including newly created strains, are described in Table 2.2.

2.2.2 Growth conditions and culture media

Fungal strains were routinely grown on Yeast Extract Peptone Dextrose media (YEPD) (10 g/L yeast extract [Sigma-Aldrich], 20 g/L Bacteriological Peptone [Oxoid] and 20 g/L dextrose [Sigma-Aldrich], at pH 5.5, (15 g/L Bacto™ Agar [Difco, Sparks, Maryland, USA] was used when making solid media), at 30 or 37 °C in a static incubator (Gallenkamp, Leicester, UK) for agar plates, and a shaking incubator (New Brunswick Scientific, Edison, New Jersey, USA) set at 200 rpm for broth cultures. Nourseothricin resistant strains were grown on 100 µg/ml nourseothricin for *C. dubliniensis* derived strains and 200 µg/ml nourseothricin for *C. albicans* derived strains (cloNAT, Werner Bioagents, Germany). YEP-Galactose was used as an alternative carbon source media for some experiments. This was made up similarly to YEPD, however 20 g/L dextrose was replaced by 20 g/L galactose (Sigma-Aldrich). Yeast Nitrogen Base (YNB) (YNB 6.7 g/L [Sigma-Aldrich], dextrose 10 g/L [Sigma-Aldrich]) was used for growing fungal strains in nutrient poor conditions, or when using certain other drugs or reagents. YNB with Amino Acids and Ammonium Sulphate (Sigma-Aldrich) and YNB Without Amino Acids and Ammonium Sulphate (Sigma-Aldrich) were both used. Spider medium was composed of nutrient broth 20 g/L (BD), mannitol 20 g/L (Sigma-Aldrich) and K_2HPO_4 4g/L (Sigma-Aldrich) with the pH adjusted to 7.2 with NaOH (Sigma-Aldrich). All strains were grown aerobically unless otherwise stated.

Escherichia coli strains were cultured in Lysogeny broth (L broth) or L agar, containing either 100 µg/ml ampicillin (Penbritin, GlaxoSmithKline, Rathfarnham, Dublin) or 50 µg/ml chloramphenicol or kanamycin at 37 °C.

Strain	Genotype	Parent	Reference
Candida dubliniensis			
Wü284	Wild-type	/	(Morschhauser *et al.*, 1999)
Δtlo1/Δtlo2	*Δtlo1Δ/Δtlo1, Δtlo2/Δtlo2*	Wü284	(Haran *et al.*, 2014)
CaTLOβ2	*Wü284::$P_{TLOα1}$CaTLOβ2*	Wü284	(Flanagan *et al.*, 2018)
P_{ACT1}CaTLOβ2	*Wü284:: P_{ACT1}CaTLO2*	Wü284	(Flanagan *et al.*, 2018)
CaTLOγ11	*Wü284::$P_{TLOγ11}$CaTLOγ11*	Wü284	(Flanagan *et al.*, 2018)
P_{ACT1}CaTLOγ11	*Wü284:: P_{ACT1}CaTLOγ11*	Wü284	This study
CaTLOα12	*Wü284:: $P_{TLOα12}$CaTLOα12*	Wü284	(Flanagan *et al.*, 2018)
P_{ACT1}CaTLOα12	*Wü284:: P_{ACT1}CaTLOα12*	Wü284	This study
CdTLO1-HA	*Wü284::CdTLO1-3xHA*	Wü284	(Haran *et al.*, 2014)
CdTLO2-HA	*Wü284::CdTLO2-3xHA*	Wü284	(Haran *et al.*, 2014)
Candida albicans			
SC5314	Wild-type	/	(Gillum *et al.*, 1984)
ΔΔmed3	*Δmed3/Δmed3*	SC5314	(Zhang *et al.*, 2012)
AHY940	*Δleu2/LEU2*	SC5314	(Nguyen *et al.*, 2017)
ΔΔtlo	All *TLOs* deleted	AHY940	This study
ΔΔtlo::P_{TET}TLOα1	*ΔΔtlo*::P_{TET1}TLOα1-3xHA	*ΔΔtlo*	This study
ΔΔtlo::P_{TET}TLOβ2	*ΔΔtlo*::P_{TET1}TLOβ2-3xHA	*ΔΔtlo*	This study
ΔΔtlo::P_{TET}TLOγ11	*ΔΔtlo*::P_{TET1}TLOγ11-3xHA	*ΔΔtlo*	This study
ΔΔtlo::P_{ENO}TLOα1	*ΔΔtlo*::P_{ENO1}TLOα1-3xHA	*ΔΔtlo*	This study
ΔΔtlo::P_{ENO}TLOβ2	*ΔΔtlo*::P_{ENO1}TLOβ2-3xHA	*ΔΔtlo*	This study
ΔΔtlo::P_{ENO}TLOγ11	*ΔΔtlo*:: P_{ENO1}TLOγ11-3xHA	*ΔΔtlo*	This study
Saccharomyces cerevisiae			
Y152			(Eguez *et al.*, 2004)
Escherichia coli			
DH5α			New England Biolabs

Table 2.2 Strains used in this work

2.2.3 Storage of strains at -80 °C

Samples to be stored long term were kept in a -80 °C freezer, stored using a Microbank™ storage system (Pro-Lab Diagnostics, Birkenhead, Wirral, UK). This system contains a set of porous beads in a cryopreservative solution. Inoculation of the cryopreservative was carried out as described by the manufacturer.

When reactivating a strain, the vial of beads was taken out of the freezer and allowed to thaw on ice. A single bead was then taken from the vial and placed on a fresh YEPD agar plate. The plate was then streaked as normal. Vials were kept on ice and returned to the freezer as soon as possible.

2.2.4 Cell counting – haemocytometer method

To determine cell counts for yeast growing in liquid cultures, 1 ml of the culture was spun down at 14,000 rpm for 3 min, washed twice with 1 ml of PBS and then resuspended in 1 ml PBS. Cultures were diluted, 10 µl in 990 µl PBS. 10 µl of the cell suspension was added to the counting chamber of a haemocytometer (Marienfeld Superior, Germany). A Nikon Eclipse E600 light microscope (Nikon, Minato, Tokyo, Japan) at X400 magnification was used to count yeast cells in the central grid of the haemocytometer. When counting each sample, five separate, non-touching squares were tallied. These cell counts could then be used to standardise cell cultures to specific concentrations required for various procedures.

2.2.5 Cell counting – spectrophotometric method

Overnight cultures were grown in YEPD in a 37 °C shaking incubator at 200 rpm. The following day, 1 ml of these cultures was pelleted at 14,000 rpm for 1 min, washed in 1 ml PBS and finally resuspended in 1 ml PBS. The OD_{600} of a 100 µl cell solution in 900 µl PBS dilution was measured. This measurement allowed for the determination of the number of cells per ml, with an OD_{600} of 1 being the equivalent of 3×10^7 cells/ml, which was then used to standardise cell counts for experimentation.

2.2.6 Comparison of cell counting methods

To determine which cell quantification method was most appropriate, spot plate assays were used to investigate the reproducibility of each method. It was seen that for the more pseudohyphal strains, using the spectrophotometric method of cell counting, rather than the haemocytometer method, produced results with less variation when performed in triplicate. It was decided that using the spectrophotometric method was most appropriate for any experiments which required standardising cell concentrations.

2.2.7 Microscopy and cell imaging

Wet mounts of cells were prepared by placing 10 µl of cell culture to be examined on a clean microscope slide and covering with a cover slip. For observing cellular morphologies and for imaging cells, a Zeiss microscope (Carl Zeiss AG, Oberkochen, Germany) was used. Micron Optical capture software (Micron Optical, Enniscorthy, Co. Wexford, Ireland) was used to capture images and insert size bars onto images. When examining flat bottom 6, 12, 24 or 96 well plates, a Nikon TMS inverted light microscope (Nikon) was used.

2.2.8 Calcofluor White staining

Fluorescent staining of cells was performed similarly to assembling wet mounts of cultures. 10 µl of microbial culture was dropped onto a clean microscope slide, with the addition of 10 µl Calcofluor White (CFW) Stain (Sigma-Aldrich) and 10 µl of 10% (w/v) Potassium Hydroxide (Sigma-Aldrich). Cells were then viewed on a Zeiss microscope using a UV filter.

2.3 DNA Techniques

2.3.1 Routine genomic DNA extraction from yeast cells

Overnight cultures were grown in 4 ml YEPD broth in a 200 rpm shaking incubator at 37 °C. These cultures were spun down at 12,000 rpm for 2 min, and the pellet resuspended in 200 µl of breaking buffer. This suspension was transferred into a 2 ml screw cap Sarstedt tube (North Rhine-Westphalia, Nümbrecht, Germany) containing 0.3 g of glass beads, 200 µl of phenol:chloroform:isoamyl alcohol (24:24:1) (Sigma-Aldrich) was added. The tubes were then placed in a bead beater for 30 s, after which the contents were spun down at 14,000 rpm for 10 min. The aqueous phase was removed and placed in a fresh 1.5 ml Eppendorf tube (Eppendorf, Hamburg, Germany). To this tube an equal volume of chloroform:isoamyl alcohol (24:1) (Sigma-Aldrich) was added. The tube was centrifuged again 14,000 rpm for 2 min in an Eppendorf 5430 benchtop centrifuge (rotor FA-45-30-11) (Hamburg, Germany). The aqueous phase was removed and placed in another fresh tube. To this tube, 20 µl of 3 M sodium acetate (Sigma-Aldrich) and 400 µl ice cold 70% (v/v) ethanol (Sigma-Aldrich) was added, and the tube spun down for another 10 min at 14,000 rpm speed. The pellet was washed in 70% (v/v) ethanol, and finally resuspended in 35 µl water. Before the final resuspension, it was ensured that all traces of ethanol were removed from the pellet by allowing it to air dry. The concentration of the DNA was determined (described in Section 2.3.3), and 100 ng/µl stock solutions were made up for use in routine PCR (see Section 2.3.6).

2.3.2 Plasmid extraction from bacterial cells

Plasmids were extracted from *E. coli* strains using the PureYield™ Plasmid Miniprep System (Promega). The manufacturer's instructions were followed, with the exception that the purified plasmid DNA was resuspended in 30 µl molecular biology grade water (Sigma-Aldrich). Plasmid concentrations were determined as other nucleic acids (described in Section 2.3.3). Plasmids used in this work and their details can be found in Table 2.3.

Name	Description	Reference
CRISPR Plasmids		
pADH110	Plasmid for amplifying the A fragment for Hernday CRISPR protocol	(Nguyen *et al.*, 2017)
pADH119	Plasmid for amplifying a unique B fragment for Hernday CRISPR protocol using a custom gRNA oligonucleotide	(Nguyen *et al.*, 2017)
pADH137	Plasmid digested with Mss*I* to create the Cas9 cassette for Hernday CRISPR protocol	(Nguyen *et al.*, 2017)
Cloning Plasmids		
pCDRI	*E. coli* plasmid for inserting genes into *Candida* species at the *CDR1* locus. *SAT1* resistance marker. *Ampicillin* resistance.	(Moran *et al.*, 2007)
pGM161	Derivative of pCDRI for inserting genes into *C. dubliniensis* under the *ACT1* promoter	(Flanagan *et al.*, 2018)
pBluescript	*E. coli* plasmid for inserting genes into *Candida* species. *LacZ* gene disrupted by insertion into MCS, allows for blue/white screening of possible transformants on X-Gal/IPTG media. Ampicillin resistance.	(Short *et al.*, 1988)
pSFS2a	*E. coli* plasmid for inserting genes into *Candida* species. Flippase under the control of the pMAL2 promoter, allows for recycling of the cassette. Ampicillin resistance.	(Reuss *et al.*, 2004)
pJESS	Derivative of pSFS2a with 300 bp of *TLO34* 3' homology flanked by *Sac*II (5') and *Sac*I (3') restriction sites	This study
pJESS-ENO1/2/11	Derivative of pJESS with either *TLOα1, TLOβ2* or *TLO11* with a 3X-HA tag with a region of *TLO34* 5' homology following, under the control of the *ENO1* promoter flanked by *Kpn*I and *Sac*I restriction sites	This study
pNIM1	*E. coli* plasmid containing a *CaGFP* ORF under the control of a tetracycline inducible promoter (*TET1*) flanked by *Sal*I and *Bgl*II restriction sites. Contains the selection marker Ca*SAT1* conferring resistance to nourseothricin. Entire excisable cassette flanked by regions of homology for the CaADH1 locus, with restriction sites *Sac*II and *Kpn*I	(Park and Morschhauser, 2005)
pNIM1-TLO1/2/11	Derivative of pNIM1 with either *TLOα1, TLOβ2* or *TLO11* tagged with 3X-HA tag replacing the *Ca*GFP ORF	This study

Table 2.3 Plasmids used in this work

2.3.3 Quantification of nucleic acid concentrations and quality checking

After DNA or RNA was purified, quantification and quality assessment was performed using a Nanodrop spectrophotometer (Thermo Fisher Scientific) and Nanodrop 2000c software (Thermo Fisher Scientific) as per the manufacturer's instructions. This spectrophotometer measures the concentration of nucleic acid in solution in ng/µl, and also provides the 260/280 nm ratio and the 260/230 nm ratio for the sample, which measure protein and phenol contamination or carbohydrate and phenol contamination respectively. A 260/280 nm ratio of around 1.8 is deemed "pure' for DNA, or around 2 for RNA. A 260/230 nm ratio of 2.0 – 2.2 is deemed "pure", as per the manufacturer's instructions.

2.3.4 Agarose gel electrophoresis

Agarose gels were routinely prepared using 1% (w/v) Agarose (Sigma-Aldrich) in 0.5X TBE buffer. When casting a gel, 1.5 µl Gel Red (Biotium, CA, USA) was added to the tray (Thermo Fisher Scientific) before the molten agarose for a small gel (10-12 wells), or 3 µl for a larger gel (20-25 wells). Samples from routine PCR reactions (see 2.3.6) did not require mixing with loading dye as the Green buffer component of the GoTaq® system contains its own loading dye. Gels were run at 100 V using a power supply unit (Mason Technologies, Dublin, Ireland) for up to 2 h depending on the resolution and spacing required. After adequate running, the gels were visualised and imaged using an Alpha Imager Mini (Alpha-Innotech, La Chaux-de-Fonds, Switzerland) under UV light.

2.3.5 Restriction digestion and enzymatic ligation of DNA

Restriction enzymes were purchased from either Promega or New England Biolabs (NEB). The manufacturer's instructions were followed for these digestions. Double digestions were carried out in the most suitable buffer determined by the NEB or Promega calculators, available on the manufacturer's website. DNA T4 ligase was purchased from Promega. Ligations were carried out as per the manufacturer's instructions. Sizes and presence of products of digestions or ligations were checked on a 1% (w/v) Agarose gel as described in Section 2.3.4).

2.3.6 Routine polymerase chain reaction

The GoTaq® PCR Core System (Promega) was used for routine PCR. The majority of PCRs were carried out according to the following ratios; 30.75 µl H_2O, 10 µl 5X Green buffer, 5 µl $MgCl_2$ (25 mM), 1µl dNTPs (10 mM), 1 µl of each oligonucleotide (10 mM), 0.25 µl Taq polymerase, and 1 µl template DNA (1 ng/µl). DNA was quantified using a Nanodrop 2000 (Thermo Fisher Scientific). Master mixes were used for setting up multiple reactions to reduce margins for error.
Standard cycling settings were as follows; initial denaturation at 95 °C for 1 min, denaturation at 95 °C for 30 seconds, annealing at varying temperatures depending on primers for 30 seconds, extension at 72 °C for a varying amount of time (1 min per 1kb of product), returning to denaturation step for 35 cycles, and a post cycling extension step at 72 °C for 10 min. PCR products were held at 16 °C until being stored in either 4 °C or – 20 °C depending on needs. Product presence and sizes were checked on 1% (w/v) Agarose gels (see Section 2.3.4).

2.3.7 PCR clean up and gel extraction

If PCR products were to be used in further downstream applications, they were cleaned up to remove any impurities from the PCR process. This was performed using the GenElute™ PCR Clean-Up Kit (Sigma) according to the manufacturer's instructions. Cleaned PCR products were finally resuspended in 30 µl water. Product concentration and quality was assessed as in Section 2.3.3. Specific restriction endonuclease digestion products were cleaned up post electrophoresis using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions, again with the final resuspension of the product in 30 µl water.

2.3.8 Transformation of Candida *species by electroporation*

This method is adapted from (Staib *et al.*, 2001). A single *Candida* colony from a 24-48 h old YEPD plate was used to inoculate 50 ml YEPD broth in an Erlenmeyer flask and incubated overnight at 37 °C in a 200 rpm shaking incubator. The next morning, 10 ml of this culture was transferred into 40 ml fresh YEPD and the culture grown to an OD_{600} = 1.8-2.0. At this point the culture was transferred to a 50 ml polystyrene tube and centrifuged at 2,000 rpm for 5 min in a bench top centrifuge (Universal 320 [rotor

1624], Hettich Instruments, Tuttlingen, Germany). The supernatant was discarded and the pellet resuspended in 10 ml 1X TE buffer with 100 mM Lithium Acetate. This solution was incubated for 1 h at 37 °C in a 200 rpm shaking incubator. Dithiothreitol (100 mM) was added to the cell suspension and incubated for a further 30 min. After this time, the volume in the tube was made up to 50 ml with water and spun down as before. The supernatant was discarded and the pellet was washed with 50 ml ice cold water and centrifuged as before, the supernatant was discarded and the pellet resuspended in 50 µl ice cold 1 M Sorbitol. Following this, 40 µl of this cell slurry was mixed with the DNA to be transformed into the cells, with the goal to use between 500-1000 ng DNA in a volume of less than 10 µl. As a negative control, the same volume of molecular biology grade water was mixed into another aliquot of the cell slurry. This mixture was transferred to a 0.2 cm electroporation cuvette (Bio-Rad) and subjected to electroporation at 1.5 kV for 5 ms in a Bio-Rad MicroPulser™ electroporator (Bio-Rad Laboratories, Hercules, California, USA). After electroporation, 800 µl YEPD broth was added to the cuvette to resuspend the slurry. This was transferred to a 1.5 ml Eppendorf tube and at 37 °C in a 200 rpm shaking incubator for 2 h to recover. After this time, 100 µl volumes were plated on YEPD plates supplemented with nourseothricin and incubated for 24-48 h at 37 °C in a static incubator. Selection was achieved using 100 µg/ml nourseothricin for *C. dubliniensis* strains and 200 µg/ml for *C. albicans* strains. Resistant colonies were subcultured onto YEPD plates containing nourseothricin. Transformants were screened using DNA extractions performed as in Section 2.3.1 and this DNA subjected to routine PCR as in Section 2.3.6 to check for insertion of the transformation cassettes of to identify desired deletions.

2.3.9 Transformation of E. coli *strains by heat shock*

Competent *E. coli* DH5α derivative cells were purchased from New England Biolabs. These cells were stored at -80 °C prior to use. Before transformation, cells were thawed on ice. Aliquots containing 100 µl of competent cells were used for transformation, and it was to the thawed cells that DNA or ligation mix was added. Cells were incubated in a 42 °C water bath for 1 min, and then held on ice for 2 min. 1 ml of SOC outgrowth medium, supplied with the competent cells, was added and the mixture placed in a 37 °C shaking incubator for 1 h. After incubation 100 µl samples of the culture was spread on appropriate selective agar plates and incubated overnight at 37 °C.

2.3.10 Sanger sequencing and basic DNA sequence analysis

For sequences of lengths up to around 1000 bp, samples were sent to Source Bioscience (Tramore, Co. Waterford, Ireland), with the relevant amplification primers, for Sanger sequencing to be performed. Sequence files were returned in .seq and .abi file formats. DNA .seq sequence files were viewed, analysed and manipulated using Serial Cloner (SerialBasics) software. The *Candida* Genome Database (http://www.candidagenome.org/) (Skrzypek *et al.*, 2017) was used to access the sequences of previously published and annotated genes. The reference *C. albicans* genome used was *Candida albicans* SC5314 Assembly 22.

2.3.11 Outsourcing of DNA cloning services

While some of the DNA cloning in this work was performed in house, some aspects of the cloning process were outsourced to Genewiz (South Plainfield, New Jersey, USA). These were typically large, customised fragments that were first synthesised by the company, then cloned into plasmids we had provided. Sequence files of the desired insertions were designed in Serial Cloner and submitted to Genewiz. After synthesis and cloning the company returned the plasmids along with quality assurance documentation including sequence files confirming the cloning was successful. Images of agarose gels of the plasmids which had been digested with restriction enzymes to show the correct sizes of plasmids were also sent, these were replicated in house as another confirmation step. Further detail about which experiments made use of externally prepared cloned materials is provided in the relevant chapters.

2.4 RNA Techniques

2.4.1 RNA extraction for qRT-PCR

Overnight cultures were grown in a 37 °C shaking incubator at 200 rpm. The following day, 1 ml of these cultures was pelleted at 14,000 rpm for 1 min, washed in 1 ml PBS and finally resuspended in 1 ml PBS. The OD_{600} of a 100 µl cell solution in 900 µl PBS dilution was measured, and the volume necessary to add to 25 ml of YEPD broth in order to achieve a starting OD_{600} of 0.1 was determined. These volumes were added and the flasks placed back in a 37 °C shaking incubator at 200 rpm until the OD_{600} reached 0.8, this was usually after 4-5 h. Once the desired OD_{600} was reached, an RNeasy Mini Kit (Qiagen) was used to extract RNA from the culture using mechanical disruption to lyse the cells.

2.4.2 cDNA synthesis for qRT-PCR

RNA used for cDNA synthesis was extracted from cells using an RNeasy Mini Kit (Qiagen) as described above. Following extraction, the RNA was treated with TURBO DNase (Thermo Fisher Scientific), to clean up the samples and remove any contaminating DNA.

The concentration of RNA in ng/µl in samples was then determined using a Nanodrop 2000 (see Section 2.3.3). RNA concentrations were standardised, and 11 µl solutions were made up at 100 ng/µl. In cases where samples contained less than 100 ng/µl RNA, the concentration of the weakest sample was used as the point of standardisation, and the other samples were made to match this in 11 µl volumes.

cDNA synthesis was carried out using the Superscript III Reverse Transcriptase kit (Invitrogen). Using standardised concentrations of RNA, 10 µl RNA was combined with 1 µl dNTPs (Promega) and 1 µl oligo$(dT)_{15}$ primer (Promega), and this was incubated in a SimpliAmp thermocycler (Applied Biosystems) at 65 °C for 5 min. After incubation, 4 µl First Strand buffer was added to the RNA samples, as well as 1 µl Superscript III enzyme, 1 µl DTT, 1 µl RNase Out and 1 µl H_2O. The samples were incubated at 50 °C for 60 min and then 70 °C for 15 min, before being held at 4 °C for storage. If the cDNA was

to be kept long term, it was stored at -20 °C. Where possible, master mixes were used to minimise error.

2.4.3 Quantitative Real-Time PCR (qRT PCR)

For each sample, 7.5 µl SYBR Green (Applied Biosystems), 0.375 µl of each of the forward and reverse primers and 5.75 µl H_2O was added to 1 µl cDNA, prior to use cDNA was diluted 1:10 in H_2O. Primers used in qRT PCR are detailed in Table 2.1.

PCR reactions were carried out in an AB7500 Real Time PCR machine (Applied Biosystems) at the default settings for SYBR reactions. For each biological replicate, three technical replicates were run. The *ACT1* gene was used as an endogenous control to determine C_T values, and the comparative C_T method (Schmittgen and Livak, 2008) was used to analyse the data. Initial data was exported to Microsoft Excel and the C_T values determined, further analysis was then performed and graphically represented using GraphPad Prism ver. 9 (San Diego, California, USA).

2.5 Protein Techniques

2.5.1 Crude total protein extraction from yeast cells

Overnight yeast cultures were incubated at 37 °C in a shaking incubator at 200 rpm in 4 ml of the required broth medium (e.g. YNB, YEPD and Spider medium). After overnight incubation, the cultures were diluted 1 in 4, returned to the incubator and incubated until the OD_{600} measured between 1-2. At this point, 2 ml of culture was harvested in a 1.5 ml Eppendorf tube and centrifuged at 14,000 rpm for 3 min. The supernatant was discarded and the pellet resuspended in 1 ml H_2O. Next, 150 µl YEX-Lysis Buffer (1.85 M NaOH, 7.5% (v/v) β-mercaptoethanol) was then added and the cells incubated on ice for 10 min, before the addition of 150 µl of ice-cold 50% (v/v) Trichloroacetic Acid and a further incubation on ice for a further 10 min. Cells were pelleted at 14,000 rpm at 4 °C for 5 min, the supernatant was discarded and the pellet was resuspended in 100 µl Sample Buffer (40 mM Tris/HCl pH 6.8, 8 M Urea, 5% (w/v) SDS, 100 mM EDTA, 1% (v/v) β-mercaptoethanol, 0.1 g/L bromophenol blue [BioRad]) and neutralised with 15 µl unbuffered 1 M Tris. Samples were incubated at 37 °C for 15 min and pelleted at 14,000 rpm for 10 min at room temperature. Supernatant was transferred to new Eppendorf tubes and stored at -20 °C.

2.5.2 SDS-PAGE for Western Blot analysis

SDS-PAGE was performed using Bio-Rad (Bio-Rad Laboratories) precast gels, with 10 µl of protein sample being loaded in each well, and 5 µl EZ-Run™ Pre-Stained *Rec* Protein Ladder (Fisher BioReagents) running alongside. Any empty wells were filled with blank sample buffer. Gels were run in a 1X electrode running buffer (25 mM Tris, 192 mM Glycine, 1% (w/v) SDS) at 150 V until the dye front had migrated to the bottom of the gel.

2.5.3 Protein transfer to PVDF membrane for Western Blot

After SDS-PAGE, the gel was soaked in Twobin Buffer (25 mM Tris, 192 mM Glycine, 20% (v/v) Methanol, 0.004% (w/v) SDS). Immobilon PDVF membrane (Sigma) was soaked in 100% Methanol for 20 seconds, transferred to sterile distilled water for 2 min

and then soaked in Twobin buffer for 5 min. The gel sandwich for transfer was constructed as follows (negative to positive); mesh, damp 3 MM Whatman paper, gel, PVDF membrane, damp 3 MM Whatman paper, mesh. The cassette was closed and placed in the tank and submerged in Twobin buffer. The transfer was carried out at 300 mA for 1 h. A magnetic stir bar was used to circulate the buffer, and an ice block was used to cool the buffer and prevent overheating. Successful transfer was assumed if protein markers from the ladder had visibly transferred onto the membrane.

2.5.4 Immunoblotting membranes for Western Blot

Membranes were covered in a Blocking Solution (5% Blotting-Grade Blocker [Bio-Rad] in 1X PBS, 0.1% (v/v) Tween 20 [Bio-Rad], 0.005% (v/v) Antifoam Y-30 Emulsion [Sigma]) for 1 h with agitation on a shaking table at room temperature. Blocking solution was replaced with Primary Antibody Solution (12CA5 Antibody [Roche Diagnostics] diluted 1:1500 Blocking Solution) and incubated again with agitation for 45 min. The membrane was washed three times for 10 min each with 1X PBS with 0.1% (v/v) Tween 20 then the membrane was covered with Secondary Antibody Solution (Amersham™ ECL™ Anti-Mouse IgG [Fisher Scientific] diluted 1:2000 in Blocking Solution) and incubated with agitation for 1 h. The membrane was then washed 3 times for 5 min and once for 10 min with 1X PBS with 0.1% (v/v) Tween 20.

2.5.6 Development and Imaging

The membrane was developed using Pierce™ ECL Western Blotting Substrate (using 750 µl of each solution) and allowed to develop for 5 min before imaging on an ChemiDoc Imaging System (Bio-Rad). Initial exposure for 1 min was performed and imaged, and then further, longer exposures were performed if necessary.

2.6 Phenotypic analysis

2.6.1 Growth rate analysis

Overnight cultures of each strain to be examined were prepared in 4 ml YEPD broth at 37 °C in a 200 rpm shaking incubator. From this overnight culture, 1 ml was placed in a sterile 1.5 ml Eppendorf tube and spun down at 14,000 rpm for 1 min. The pellet was washed in 1 ml PBS and then resuspended in 1 ml PBS. The suspension was diluted, 70 µl of cells in 630 µl PBS and the OD_{600} of this dilution was determined. This was then used to determine how much of the resuspended cells were to be added to 25 ml of YEPD broth, or YEP-Gal broth, in a screw top Erlenmeyer flask to give a starting OD_{600} of 0.1. Flasks were then placed in a shaking incubator at 200 rpm at 37 °C and the OD_{600} measured every 2 h for 8 h. GraphPad Prism was used to analyse the data, the Exponential Growth Rate equation was used to determine the growth rate using time points 2 h to 6 h for the calculations.

2.6.2 Hyphal induction in static liquid media

Overnight cultures of each strain to be examined were prepared in 4 ml YEPD broth at 30 °C in a 200 rpm shaking incubator. Of this culture, 1 ml was spun down at 14,000 rpm for 3 min, washed twice with 1 ml of PBS and then resuspended in 1 ml PBS. The suspension was diluted, 70 µl of cells in 630 µl PBS and the OD_{600} of this dilution was determined. The volume needed to add to 2 ml of YEPD broth with 10% (v/v) Foetal Calf Serum (FCS) to reach $1x10^6$ cells/ml was determined, and these volumes were each added to 2 ml of YEPD with 10% (v/v) FCS in a 6 well plate, a separate plate was inoculated for each time point. Plates were placed in a static incubator at 37 °C. At each time point, one plate was removed from the incubator and viewed using a Nikon TMS inverted light microscope (Nikon). Samples (10 µl) of the culture were taken, placed on a clean microscope slide and subjected to Calcofluor White staining, then imaged using a Zeiss light microscope with a UV filter. This was done at 2, 4 and 6 h time points.

2.6.3 Hyphal formation on solid media

For each strain to be examined, a single colony was taken from a 24-48 h old plate and streaked in a straight line across an agar plate. These plates were incubated for 3-5 days at either 37 °C or 30 °C. After incubation, plates were imaged using a Flash and Go plate visualiser (IUL Instruments, Barcelona, Spain) connected to a PC and results were recorded.

A similar method was also used where 5 µl drops of a $1x10^6$ cfu/ml solution of each strain was spotted onto an agar plate and incubated as above. Cell suspensions were prepared according to the spectrophotometric cell counting method outlined in Section 2.2.5 and imaged as above.

2.6.4 Chlamydospore formation assay

To induce chlamydospore formation, cells were streaked lightly on to Corn Meal Agar supplemented with 1% (v/v) Tween 80 (Sigma). Streaks were covered with a glass coverslip and then incubated at 22 °C in the dark for 5-7 days, then imaged on a Zeiss light microscope.

2.6.5 Spot plate assays

For the preparation of spot plate assays, serial dilutions of $1x10^6$, $1x10^5$, $1x10^4$, $1x10^3$, $1x10^2$ and $1x10^1$ cells/ml were prepared. The initial suspension of $1x10^6$ cells/ml was obtained using the spectrophotometric cell counting and dilution method as outlined in Section 2.2.5. These dilutions were then plated on agar plates in volumes of 5 µl using a multi-channel pipette. The agar plates were prepared by the additions of varying concentrations of either oxidative stress inducers, cell wall disrupting chemicals, antifungals or other solutions to molten YEPD or YNB agar. These plates were incubated at 37 °C in a static incubator for 2-3 days. Images of the plates were taken using a Flash and Go plate visualiser connected to a PC.

2.6.6 Minimum inhibitory concentration assays

A 96 well plate was set up such that columns 2 – 11 contained a 2-fold dilution series of a specific stressor, with each well containing 100 µl of media. The first and last columns acted as positive (no inhibitor) and negative controls (no cell inoculation), respectively. An overnight culture of the strains to be tested were grown overnight in YEPD in a 37 °C shaking incubator at 200 rpm. Cell density was quantified using the spectrophotometric method described in Section 2.2.5, and a solution of $1x10^6$ cells/ml was made up in PBS. Samples containing 100 µl of this solution were then added to each well. The plate was then incubated overnight at 37 °C in a static incubator. After 18 h the growth in each individual well was measured using a Tecan plate reader (Tecan Trading AG, Männedorf, Switzerland) measuring the absorbance at 540 nm. The data was subsequently compiled and visualised in GraphPad Prism.

2.6.7 Biofilm formation on plastic surfaces

Overnight cultures of the strains to be tested were grown in YEPD overnight at 37 °C shaking incubator at 200 rpm. The following day, 100 µl of this culture was transferred to 4 ml YNB supplemented with 100 mM glucose and this was incubated overnight again at 37 °C and 200 rpm.

Following this, cells were pelleted, washed in PBS and resuspended in 1 ml YNB with 100 mM glucose at a cell density of 10^6 cells/ml. Samples containing 100 µl of each strain were placed in triplicate in a 96-well plate and this plate was incubated at 37 °C in a static incubator for 90 min. After incubation, the medium was aspirated and the wells washed with 150µl 1X PBS. The 150µl of Spider medium was added to each well and the plates returned to the 37 °C static incubator for 24 and 48 h. Duplication of the plates was used to allow removal of one plate at 24h and the second a 48 h.

The crystal violet biofilm assay was used to determine biomass and quantify biofilm formation. After incubation, wells were washed twice with 1X PBS and then 125µl of 0.4% (w/v) Crystal Violet (Sigma-Aldrich) was added to each well and the plates stained for 10 min at room temperature. After this time, the stain was removed and the wells washed 3 times with 1X PBS. The plates were left to air dry, and then 100µl of 95% (v/v) ethanol was used to solubilise the dye. The covered plates were then left to incubate for 15 min at room temperature. Following this, the contents of the wells was mixed by

pipetting up and down, and the absorbance of the medium in each well was measured using a Tecan plate reader (Tecan Trading AG) measuring the absorbance at 540 nm. Results were used to determine levels of biofilm formation and were visualised using GraphPad Prism.

2.6.8 Galleria mellonella *infection model*

Wax moth larvae (*Galleria mellonella)* were purchased from Live Foods Direct (Sheffield, England), and later from livefoods4u (Amazon.co.uk), and stored at 15 °C in wood shavings in the dark. Larvae were used within two weeks of arrival. Larvae between 0.2 and 0.3 g were selected for infection. For each different inoculum, 10 larvae were placed in a petri dish with wood shavings and lined with Whatman filter paper. An inoculum of 1 X 10^6 cells/20 µl was injected into each larvae via a 30G insulin U-100 Micro-Fine syringe (BD). The site of injection was the last left proleg, allowing direct injection into the haemocoel. Infected larvae were then incubated at 30 °C. At 24, 48 and 72 h time points the larvae were counted to determine the number of surviving larvae. Data was analysed using GraphPad Prism and Kaplan-Meier curves generated..

Chapter 3

Expansion of the of *Candida dubliniensis TLO* gene repertoire with *TLO* genes from *Candida albicans*

3.1 Introduction

3.1.1 Candida dubliniensis

Previously misidentified as *C. albicans*, a novel species of *Candida* was identified in HIV patients with oropharyngeal candidiasis in Dublin, Ireland in 1995. The subsequently named *Candida dubliniensis* was distinguished from *C. albicans* due to differences in phenotypic and genotypic characteristics (Sullivan *et al.*, 1995). The earliest identified isolate later determined to be *C. dubliniensis* was from the UK, from the lung of a deceased patient in 1957. It is unknown, however, if *C. dubliniensis* had any role in the disease causing death. Along with *C. albicans, C. dubliniensis* is the only other member of the *Candida* spp. that forms true hyphae (Sullivan and Coleman, 1998).

Distinguishing between *C. dubliniensis* and *C. albicans* on solid conventional media is difficult, as both produce creamy white colonies on routinely used media. However, *C. dubliniensis* does not grow well, if at all, at 42 °C or higher. The *Candida* specific chromogenic medium CHROMagar can also be used to differentiate between *Candida* spp. including differentiating between *C. dubliniensis* and *C. albicans.* However, definitive means of differentiating *Candida* spp. rely on genotypic testing, such as PCR tests (Sullivan and Coleman, 1998, Abaci *et al.*, 2008).

In recent years, the distribution of *Candida* spp. causing invasive candidiasis has been changing. There has been an upward shift in the frequency of non-*albicans* spp. causing infection. However, *C. albicans* is still the most prevalent species isolated from ICU candidemia, in invasive candidiasis after solid organ transplant and it is the most common pathogen in neonatal ICU infection (Lamoth *et al.*, 2018). In the case of OPC, a review of 14 separate studies between 2008 and 2017 found that *C. albicans* was the most common cause of OPC, with frequencies of between 37.2 – 95.2%. *C. dubliniensis* was observed in OPC between 1.48 – 48.9%, being reported in 9 of the 14 studies reviewed (Patil *et al.*, 2018).

3.1.2 Comparative analysis of Candida dubliniensis *and* Candida albicans

Estimates suggest that *C. albicans* and *C. dubliniensis* diverged about 20 million years ago, and genomic comparison shows that 98.1% of the total 5569 genes in *C. dubliniensis*

are positionally conserved in *C. albicans*. On the DNA level, 96.3% of the genes share over 80% nucleotide identity (Jackson *et al.*, 2009). The *C. dubliniensis* genome is approx. 14.6 Mbp in size compared to the 14.9 Mbp size of the *C. albicans* genome. While many genes are positionally conserved between the two species, the karyotype of *C. dubliniensis* is more fragmented and contains haploid and diploid regions (Magee *et al.* 2008). The *C. dubliniensis* also contains MRS regions on all chromosomes bar one chromosome homologous to chromosome R in *C. albicans* (Jackson *et al.,* 2009). Investigations have confirmed the presence of centromere sequences in *C. dubliniensis* (Padmanabhan *et al.* 2008) however, it is unknown whether these sequences are present on all chromosomal fragments. While they are closely related genetically, many phenotypic differences exist between *C. albicans* and *C. dubliniensis.*

The morphological responses of *C. dubliniensis* to environmental conditions require stronger stimuli than *C. albicans*. This could be due to differences in the signalling pathways activating transcription necessary for these responses. Transcriptional profiling of *C. dubliniensis* in response to environmental stress has found that *C. dubliniensis* is less tolerant of thermal, oxidative and osmotic stress than *C. albicans*. An example of this is the *ENA21* gene for salt tolerance in *C. albicans* that has a homolog in *C. dubliniensis*. Increased expression of this gene can increase salt tolerance. It has been seen by transcriptional analysis that there is no significant upregulation of this gene in *C. dubliniensis* in high salt conditions. Supplementation of *C. dubliniensis* with *CaENA21* in a forward genetic screen was able to increase tolerance to osmotic stress. The *CaENA21* and the *CdENA21* genes share 91% identity, but it appears to be differences in transcript stability that affects the salt tolerance response (Enjalbert *et al.*, 2009). Differences in response to environmental conditions can also be seen in the deficiencies of *C. dubliniensis* when grown at higher temperatures (42-49 °C). The forward screening experiments by Enjalert *et al.* 2009, where the *ENA21* salt tolerance phenotype was described, were unable to identify any *C. albicans* genes that could increase heat tolerance in *C. dubliniensis,* indicating that this phenotypic difference could be the result of complex genetic control (Sullivan *et al.*, 2005, Enjalbert *et al.*, 2009). With such highly similar genomes, it would appear that differential regulation of gene expression and stress-related regulatory pathways are the most likely cause of the manifold phenotypic differences between *C. albicans* and *C. dubliniensis*.

One of the most important phenotypic differences between the two species is that *C. dubliniensis* forms hyphae less readily than *C. albicans in vivo* and *in vitro*, and given

the importance of morphogenesis in *C. albicans* pathogenesis this could be a significant contributing factor to the differential virulence of the two species. In one study, designed to dissect the factors that induce hypha formation, nutrient depletion was shown to be essential for a strong hyphal response in *C. dubliniensis*, while *C. albicans* can form hyphae in conditions with high and low nutrient availability (O'Connor *et al.*, 2010).

Regulators *RIM101, SFL2, TEC1* and *UME6* were found to be important for hyphal formation after a shift to alkaline pH in *C. dubliniensis,* while the temperature shift induced the expression of the *EFH1* and *CPH1* hyphal growth regulators. While *C. dubliniensis* lacks HAGs that are found in *C. albicans,* such as *ALS3*, *EED1*, *HYR1* and *SAP5*, it can still for hyphae under certain environmental conditions, however, it is thought that the absence of these genes could be a reason behind its reduced virulence (O' Connor *et al.,* 2010).

It is thought that the adaptive ability of *C. albicans* to switch between morphologies is a major factor in its success as a coloniser and an opportunistic pathogen (Moran *et al.*, 2012). The morphological plasticity of *C. albicans* is thought to have a role in virulence, as different morphologies have been seen to be prevalent in different stages of infection (Sudbery, 2011). Mutants that have the inability to switch forms and are trapped either in hyphal or yeast form cannot successfully infect the bloodstream in murine models. Therefore, it is thought that both forms are required for successful infection, and that the ability to switch between forms is also necessary (Noble *et al.*, 2017). The tighter control of the yeast-hyphal switching system in *C. dubliniensis* may explain why fewer systemic infections are caused by this fungus compared to the number of those caused by *C. albicans*.

The most significant differences identified between the genomes of *C. albicans* and *C. dubliniensis* have been found in the composition of gene families related to virulence, providing another possible explanation for the differential virulence of the two species. These gene families include the *IFA* family of putative transmembrane proteins, the *SAP* family of secretory aspartyl proteinases, the *ALS* family of agglutinin-like sequences and the *IFF* gene family which includes genes induced during hyphal formation in *C. albicans*. These gene families are expanded in *C. albicans* and may be either examples of widespread gene loss in *C. dubliniensis* or the families not expanding to the same extent as in *C. albicans*. Both species have a wide array of *SAP* genes, but *C. albicans* has evolved

additional genes in this family, with *C. albicans* having ten members of this family, and *C. dubliniensis* having eight (Moran *et al.*, 2012, Jackson *et al.*, 2009). Two genes in this family, *SAP5* and *SAP6*, are important, hyphal specific, virulence factors in *C. albicans* but are missing in *C. dubliniensis* (Moran *et al.*, 2012). In addition, while *C. dubliniensis* has six *ALS* genes, *C. albicans* has eight (Jackson *et al.*, 2009). One *ALS* gene absent in *C. dubliniensis* is *ALS3*, which is recognised as one of the most important virulence factors in *C. albicans* (Moran *et al.*, 2012). The *IFA* gene family in *C. dubliniensis* has 21 loci, however, 14 of these have been determined to be non-functional or fragmented, meaning only 7 functional loci remain. *C. albicans* has 25 functional *IFA* loci, and 6 non-functioning. The comparison of this gene family across the species shows both the increased expansion of gene families in *C. albicans* as well as the trend of gene loss in *C. dubliniensis* (Jackson *et al.*, 2009). The *IFF* gene family encodes a number of cell wall proteins important for hyphal formation in *C. albicans*. In *C. dubliniensis*, the *HYR1* gene is missing from this family. *HYR1* has been shown to be important in *C. albicans* in neutrophil killing, and also plays a role in biofilm formation (Moran *et al.*, 2012, Jackson *et al.*, 2009).

The greatest difference seen between the genomes of *C. albicans* and *C. dubliniensis* is in the copy number of genes in the Telomere-associated ORF (*TLO*) family, a family of putative transcription factors, which highly expanded in *C. albicans* compared to *C. dubliniensis* (Jackson *et al.*, 2009).

3.1.3 Candida dubliniensis *and the* TLO *gene family*

Candida albicans strains can have 10–15 *TLO* genes, whereas *C. dubliniensis* only has two members of this gene family. Other species of *Candida* and other yeasts only have one *TLO* gene. *C. dubliniensis* has a single gene at the ancestral locus, Cd36_35580 (*CdTLO2*) and an additional gene at Cd36_72860 (*CdTLO1*) which has evolved through a transposition event, information about these loci can be found in Table 1.1. The *C. dubliniensis TLO* repertoire is more like that of the ancestral state compared to *C. albicans* (Jackson *et al.*, 2009).

The two *C. dubliniensis TLO* paralogs share 74.9% nucleotide identity, making them relatively well diverged compared to the *TLOs* in *C. albicans* which are around 82% identical across all *TLOs* (Jackson *et al.*, 2009, Anderson *et al.*, 2015). The genes encode proteins of 320 and 355 amino acids for Tlo1 and Tlo2 respectively. With these proteins

showing 81% identity in the Med2 domain, and 50% identity in the C-terminus (Haran *et al.*, 2014).

The *TLO* gene families of *C. dubliniensis* and *C. albicans* are very different. The expansion of the family to up to 15 members, divided into different clades in *C. albicans* contrasts heavily to the two member family found in *C. dubliniensis*. Given the extent of this divergence between two otherwise highly similar species, we hypothesise that this difference in *TLO* repertoire may contribute to the phenotypic differences between the two species, and may help explain why *C. albicans* the more pathogenic of the two.

The relatively small size of the *C. dubliniensis TLO* gene family has enabled extensive work to be carried out to characterise the functions of these genes, including deletion of *CdTLO1, CdTLO2* or both. In this work, Haran *et al.* (2014) showed that deletion of the two *CdTLO* genes resulted in a strain that grew more slowly than WT *C. dubliniensis* Wü284, had defects in hyphal induction and was sensitive to oxidative stress, among other defects. Reintroduction of *CdTLO1* into this backgrounds was able to restore growth rate, hyphal production and resistance to oxidative stress. Reintroduction of *CdTLO2* was able to restore growth rate and oxidative stress resistance. *CdTLO2* was able to restore hyphal growth to some extent, but not to the same level as *CdTLO1* reintroduction (Haran *et al.*, 2014).

3.1.4 Aims of this work

The purpose of this part of the thesis was to investigate if different *C. albicans TLO* genes (from the three clades) have different functions. This involved heterologously expressing a representative *TLO* gene from each *C. albicans* clade, *CaTLOα12, CaTLOβ2 and CaTLOγ11*, in the *C. dubliniensis* Wü284 WT background. Each gene was expressed either using their native promoter or the well characterised stronger promoter *ACT1*. Due to the fact that the reference sequence for SC5314 is not complete, the *CaTLOβ2* promoter sequence was not available, so a fusion of the *CaTLOβ2* gene to the *CaTLOα1* promoter was used to give a 'native level of expression'. The effect of the heterologous expression of these genes on the phenotype of *C. dubliniensis* was then examined using a variety of tests, including virulence assays.

3.2 Materials and Methods

3.2.1 Construction of C. dubliniensis *strains expressing* C. albicans TLO *genes*

The strains used in this chapter are described in Table 2.2. Strains expressing P_{ACT1}*CaTLOγ11* and P_{ACT1}*CaTLOα12* were generated by amplifying the appropriate gene and cloning it into pGM161, see Table 2.3. A map of the insertion fragment can be found in Figure 3.1, where the enzymes used for insertion of the *TLO* genes, *Xho*I and *Hin*dIII, are highlighted. After linearisation of the plasmid with *Nco*I, clean-up of the digested fragment and quantification of DNA, the constructs were transformed into the *CDR1* locus of *C. dubliniensis* wild type strain Wü284, as described in Section 2.3.8.

3.2.2 Confirmation of mutant genotypes

The integration of the *TLO*-promoter constructs into the *CDR1* locus of *C. dubliniensis* was confirmed by routine PCR using TagR and M13R oligonucleotides which amplify a fragment of DNA at the junction of the cassette and the 3' *CDR1* sequence, see Figure 3.1. Products were visualised via agarose gel electrophoresis as described in Section 2.3.4.

3.2.3 Confirmation of TLO *mRNA expression in* C. dubliniensis *Wü284*

Expression of mRNA from the heterologous *CaTLO* genes in Wü284 was quantified by qRT-PCR analysis as described in Section 2.4, using the oligonucleotide primers specified in Table 2.1.

3.2.4 Phenotypic analysis of strains

Phenotypic analysis of the strains was performed as described in Section 2.6.

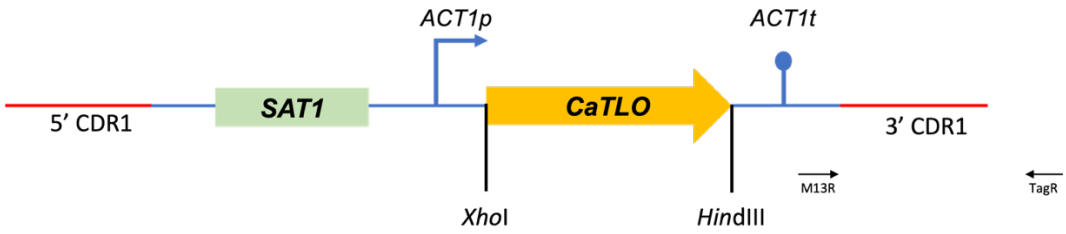


Figure 3.1 Map of cassette for integrating *CaTLO* genes into *C. dubliniensis*

This cassette in pGM161 allows the integration of genes under the control of the *ACT1* promoter into the *CDR1* locus of *C. dubliniensis.* The *SAT1* cassette contains a nourseothricin resistance marker that can be used for selection. The *Xho*I and *Hin*dIII restriction enzymes were used to insert the selected *TLO* genes into the cloning site. Black arrows indicate binding regions for M13R and TagR primers used for confirmation of integration of the cassette into the desired location in the *C. dubliniensis* genome.

3.3 Results

3.3.1 Confirmation C. dubliniensis *Wü284::CaTLO strains and qRT-PCR analysis*

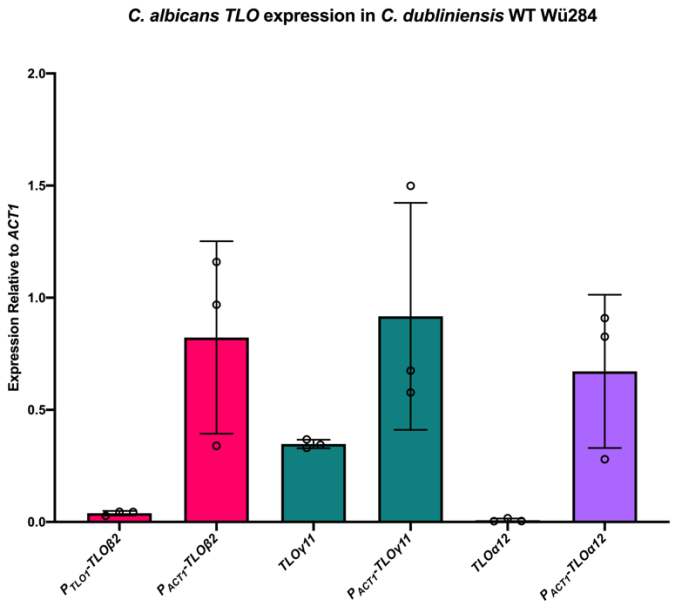
Routine PCR analysis was performed to confirm the integration of the selected *CaTLO* genes into one allele of the *CDR1* locus of wild-type *C. dubliniensis* Wü284. qRT-PCR was performed on the strains in order to examine the expression levels of each *CaTLO* in the Wü284 background. Oligonucleotide primers for each *TLO* can be found in Table 2.1. Strains tested included strains previously generated where *CaTLOα12* and *CaTLOγ11* were expressed under their native promoters, and strains where *CaTLOβ2* was expressed under the *ACT1* promoter and under the *CaTLOα1* promoter. The *CaTLOα1* promoter was used as a proxy for native expression of *CaTLOβ2* because the reference genome sequence for *C. albicans* SC5314 is incomplete, and the full sequence for the *CaTLOβ2* promoter was not available. Any reference to expression from the native promoter with regards to *CaTLOβ2* refers to expression of this gene from the *CaTLOα1* promoter.

Expression of target genes was compared to expression of the native *ACT1* gene as an endogenous control. Figure 3.2 depicts the results of this analysis and confirms that *CaTLO* genes are being expressed in the Wü284 background, and that those under the *ACT1* promoter are being expressed at a significantly higher level (0.9-0.7 x *ACT1*) than those under the control of the native promoters (0.4-0.01 x *ACT1*).

3.3.2 Colony and cellular morphology

The phenotypes of all strains expressing the selected *CaTLO* genes under the control of either the native or *ACT1* promoters were examined to investigate if overexpression of *CaTLO* genes affected the morphology of the parental *C. dubliniensis* strain. Examination of the colony morphology of *C. dubliniensis* Wü284 strains expressing *CaTLO* genes revealed that after 48 h of growth on YEPD agar at 37 °C, all strains appeared morphologically identical to the WT, except for those expressing *CaTLOβ2*, under the control of either the native or *ACT1* promoter, Figure 3.3. The *CaTLOβ2* expressing strains both displayed wrinkled cellular morphology on the YEPD solid medium. The

A



B

	TLOβ2	P_{ACT}TLOβ2	TLOγ11	P_{ACT}TLOγ11	TLOα12	P_{ACT}TLOα12
Mean	0.03930	0.8228	0.3476	0.9171	0.008143	0.6717
S.D.	0.01025	0.4291	0.01911	0.5061	0.007629	0.3419

Figure 3.2 qRT-PCR data quantifying expression of CaTLOs in Wü284

(A) Graphical representation of qRT-PCR data measuring expression of *CaTLOs* in *C. dubliniensis* WT Wü284. Each data set on the graph represents a different strain, with a *CaTLO* under either the native or *ACT1* (overexpressing) promoter. Error bars represent standard deviation, with symbols representing each of three replicates. Expression of each *TLO* was measured relative to expression of endogenous control *ACT1*. (B) Mean values and standard deviation from qRT-PCR data.

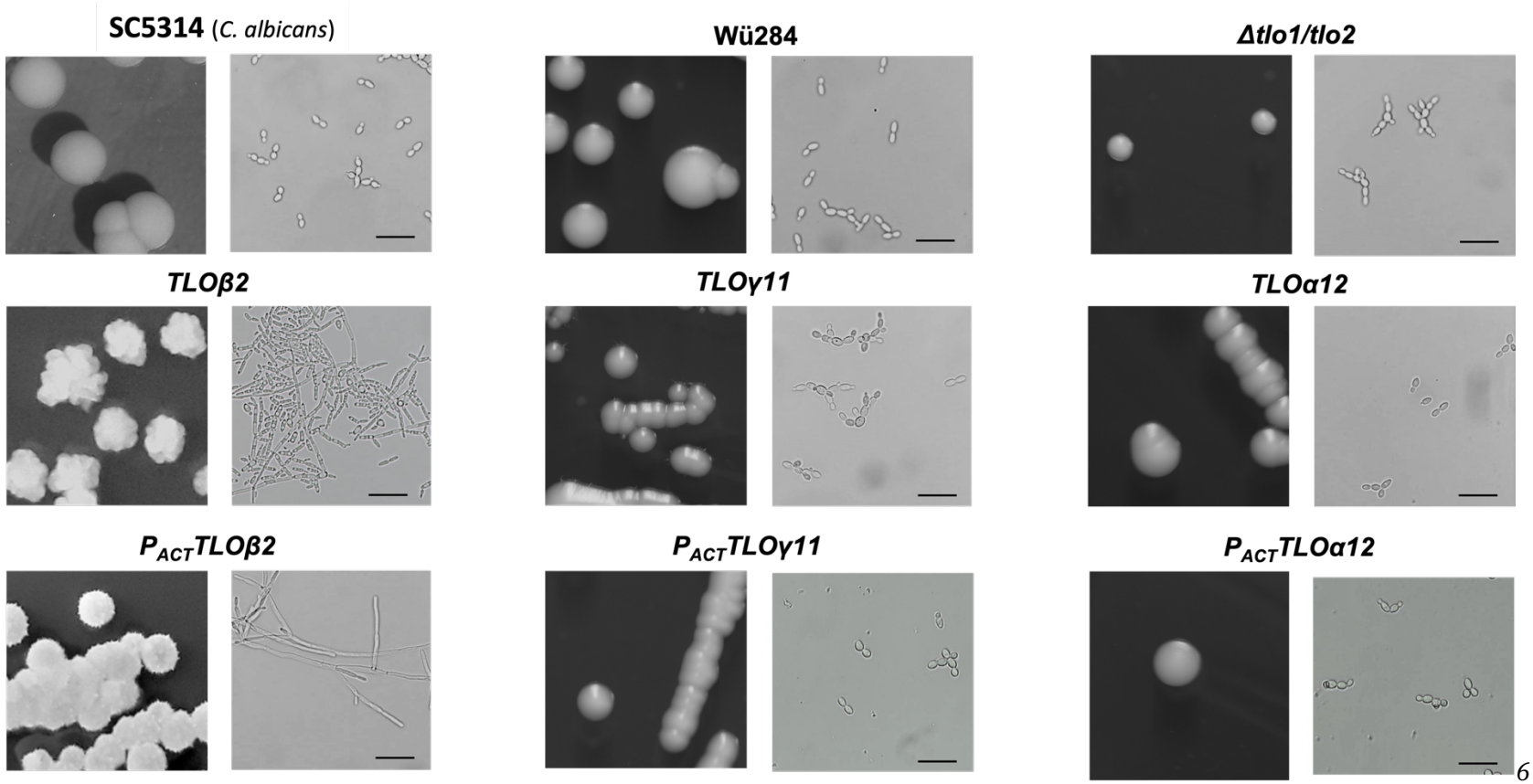


Figure 3.3 Colony and cellular morphology of *C. dubliniensis* strains expressing *CaTLOs*

Colony morphology (left) was photographed after 48h of growth on YEPD agar at 37 °C. Cellular morphology was imaged under X200 magnification after 24h of growth in YEPD liquid media. Controls included WT *C. albicans* SC5314 and *C. dubliniensis* Wü284 and the *C. dubliniensis* mutant in which both *CdTLO* genes had been deleted (*Δtlo1/Δtlo2*). Scale-bar represents 16 μm.

strain expressing *CaTLOβ2* under the ACT1 promoter appears to be more wrinkled than the strain expressing *CaTLOβ2* under the *TLOα1* native promoter.

Microscopic examination of overnight liquid YEPD cultures of the strains also showed that the strains all appear morphologically similar, growing in yeast form, apart from the strains expressing *CaTLOβ2,* which grow constitutively as hyphae*,* Figure 3.3.

3.3.3 Hyphal induction

To investigate if heterologous expression of *CaTLO* genes in *C. dubliniensis* affected the ability to form hyphae, the morphology of each construct generated was examined under a range of environmental conditions that induce hypha formation in *C. albicans.*

Incubation in YEPD + 10% FCS did not induce the formation of hyphae in any of the strains, while the *CaTLOβ2* expressing strains both maintained their hyphal form.

The morphology of the strains was also examined following growth on solid Spider medium. The morphology of *C. dubliniensis* and *C. albicans* WT strains differ greatly when grown on Spider agar, Figure 3.4. *C. albicans* SC5314 has a wrinkled colony morphology, indicating hyphal induction, whereas the *C. dubliniensis* WT strain Wü284 maintains a smooth morphology, indicating the switch between yeast and hyphal morphologies was not induced in *C. dubliniensis*. The *Δtlo1/Δtlo2* mutant strain of *C. dubliniensis* has a similar morphology to the Wü284 strain as do the *CaTLOγ11* and the *CaTLOα12* expressing strains. The *CaTLOβ2* expressing strains, however, both displayed a hyphal morphology on this medium. While the *CaTLOβ2* expressing stains of *C. dubliniensis* displayed a hyphal morphology in YEPD media, without the need for induction, there is a difference between the extent of the hyphal growth between the differently expressing strains. The strain expressing *CaTLOβ2* under the native level promoter (P_{TLO1}*TLOβ2*) appears more similar to the *C. albicans* WT SC5314, especially at 37 °C, whereas the strain expressing *CaTLOβ2* at a higher level (P_{ACT}*TLOβ2*) seems to form more extensive hyphae, with the fringes of the growth appearing to extend further into the surrounding agar, again, specifically at 37 °C.

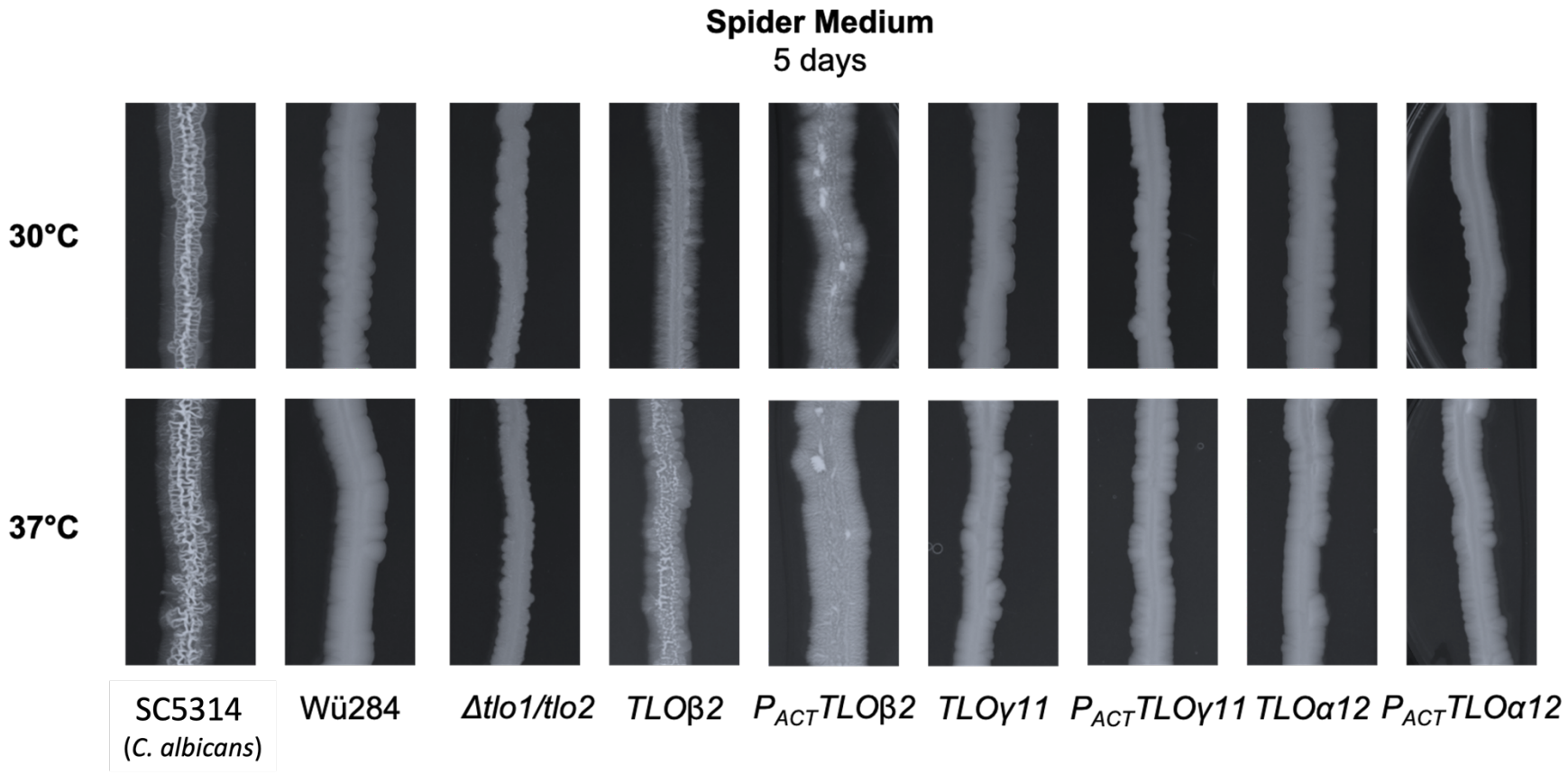


Figure 3.4 Growth of *C. dubliniensis* strains expressing *CaTLOs* on solid Spider media

The ability of strains to filament on solid Spider medium was determined by streaking a single colony of each strain onto a Spider agar plate and incubating at either 30 °C or 37 °C for 5 days. After 5 days plates were examined and photographed.

3.3.4 Chlamydospore formation

To determine if the addition of *CaTLO* genes into the *C. dubliniensis* WT background had any effect on the ability of the strains to produce chlamydospores, strains were grown on corn meal agar with 1% Tween in the dark at 22 °C for 7 days then examined under the microscope. The results of this analysis can be seen in Figure 3.5. The WT *C. dubliniensis* strain was able to form chlamydospores, as was the SC5314 strain of *C. albicans*. The *Δtlo1/Δtlo2 C. dubliniensis* strain was unable to form chlamydospores. It was seen that the addition of *CaTLOγ11* and the *CaTLOα12* had no effect on the ability of the strains to form chlamydospores under these conditions, but the strains expressing Ca*TLOβ2* were unable to form chlamydospores.

3.3.4 Growth rate analysis

Growth rate analysis was carried out in liquid YEPD media to determine if heterologous expression of *CaTLO* genes had any effect on this phenotype, Figure 3.6 (A). No significant advantage in growth was observed with the expression of any *CaTLOs* in *C. dubliniensis*. However, it was seen that the P_{ACT}*TLOβ2* strain grew significantly slower than WT Wü284, however, this may be due to the hyphal morphology seen in this strain. Statistical analysis indicated that only the *Δtlo1/Δtlo2 C. dubliniensis* strain and the P_{ACT}*TLOβ2* strain were significantly different from WT. Growth rate analysis was also carried out in YEP-galactose liquid medium, Figure 3.6 (B). These results also showed that the P_{ACT}*TLOβ2* strain grew significantly slower than WT Wü284, again possibly due to the strongly hyphal morphology. Again, only the *Δtlo1/Δtlo2 C. dubliniensis* strain and the P_{ACT}*TLOβ2* strain were statistically significantly different from WT.

Strains were grown on spot plates to determine the ability of strains to grow on YEP-Galactose in the presence of a respiration inhibitor Antimycin A, Figure 3.7, where plates are compared to a YEP-Galactose control plate. The only strain unable to grow in the presence of Antimycin A was the *Δtlo1/Δtlo2* mutant strain, indicating a role for *TLOs* in growth when respiration is inhibited. There was a distinct difference between the morphology of the *C. dubliniensis* WT strain Wü284 and the *C. albicans* WT strain SC5314, where Wü284 appears to be more sensitive to the presence of Antimycin A. It can be seen that the majority of the *C. dubliniensis* strains expressing

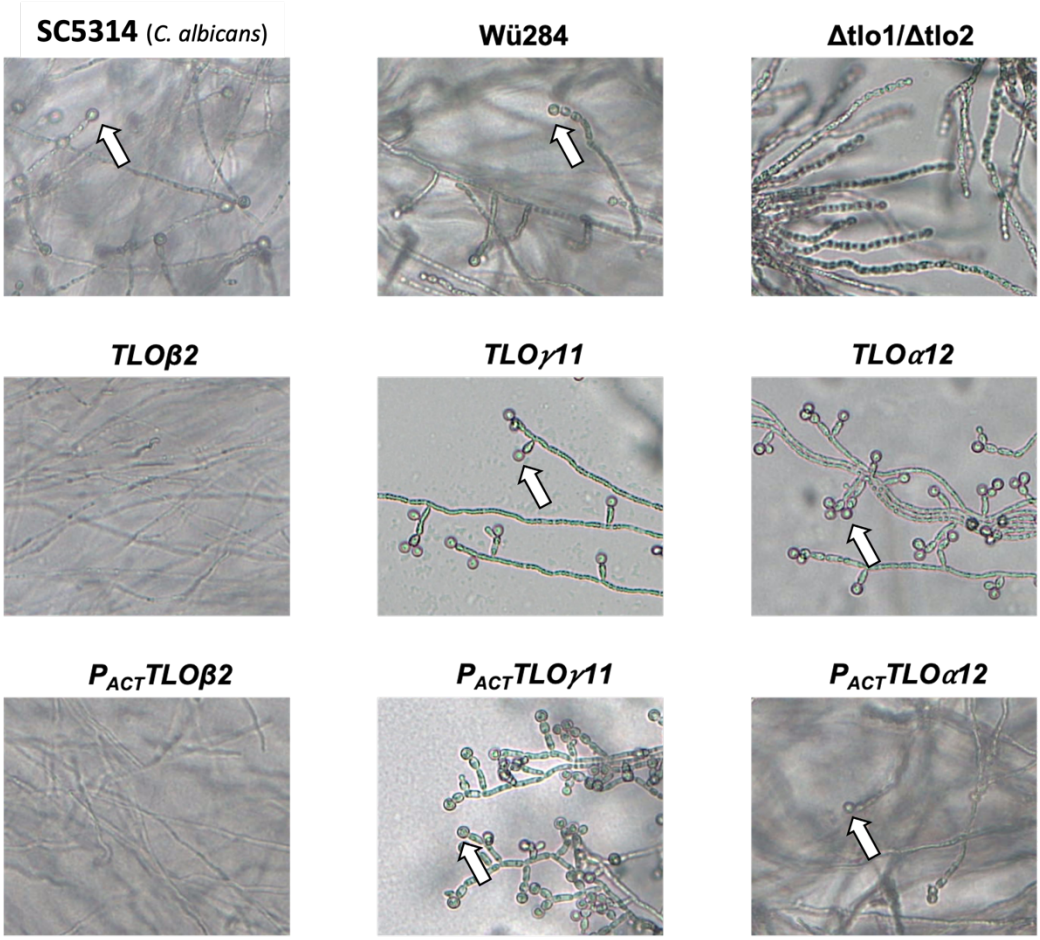
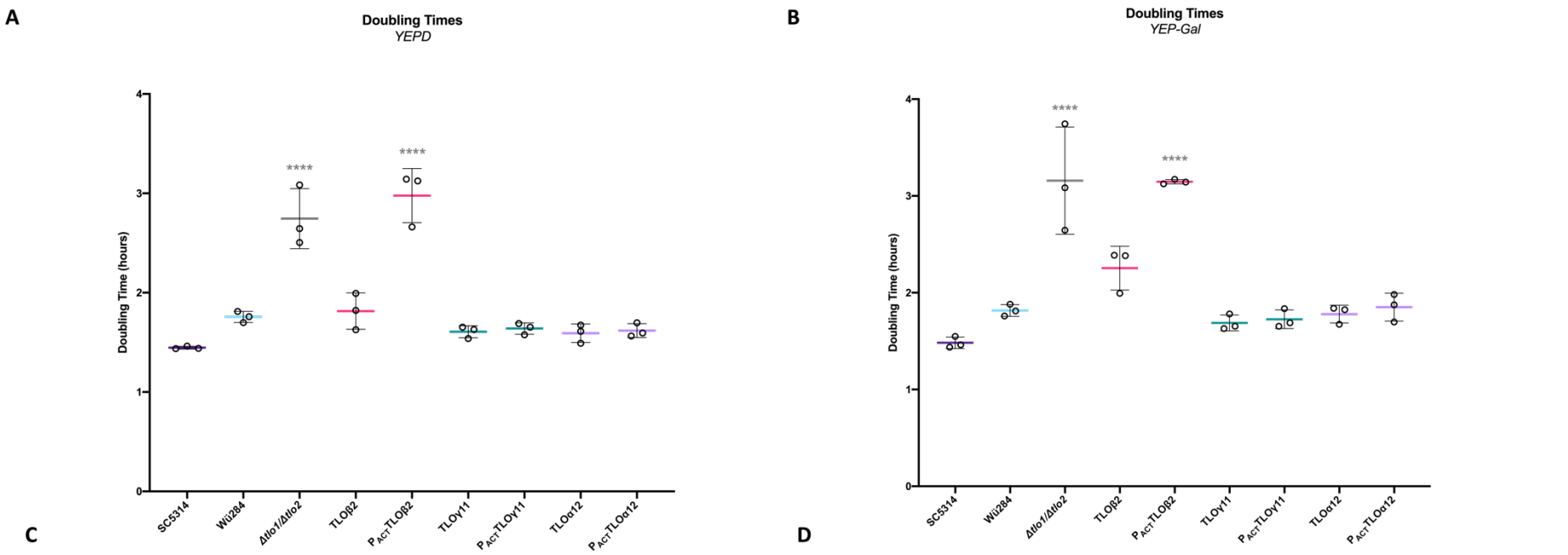


Figure 3.5 Chlamydospore formation by *C. dubliniensis* strains expressing *CaTLOs*

The ability of strains to form chlamydospores was tested by plating on corn meal agar supplemented with 1% Tween 80 and incubating for 7 days at 22 °C in the dark. The cells were imaged under a light microscope.



C

YEPD	SC5314 (*C. albicans*)	Wü284	Δtlo1/Δtlo2	TLOβ2	P_{ACT}TLOβ2	TLOγ11	P_{ACT}TLOγ11	TLOα12	P_{ACT}TLOα12
Mean (h)	1.446	1.756	2.745	1.814	2.977	1.607	1.639	1.592	1.618
S.D.	0.01365	0.05605	0.3021	0.1831	0.2727	0.06067	0.05773	0.09297	0.06976
Significantly different to Wü284	ns	-	****	ns	****	ns	ns	ns	ns
P-value	0.1364	-	<0.0001	0.9975	<0.0001	0.7847	0.9183	0.7104	0.8389

D

YEP-Gal	SC5314 (*C. albicans*)	Wü284	Δtlo1/Δtlo2	TLOβ2	P_{ACT}TLOβ2	TLOγ11	P_{ACT}TLOγ11	TLOα12	P_{ACT}TLOα12
Mean (h)	1.483	1.817	3.158	2.254	3.146	1.688	1.726	1.779	1.851
S.D.	0.05879	0.0607	0.5537	0.2263	0.02318	0.08171	0.09638	0.09149	0.1434
Significantly different to Wü284	ns	-	****	ns	****	ns	ns	ns	ns
P-value	0.3157	-	<0.0001	0.9997	0.9997	0.1138	<0.0001	0.9997	0.9997

Figure 3.6 Growth rate analysis of *C. dubliniensis* strains expressing *CaTLOs*

The doubling time of each strain is represented by a horizontal line (hours), with error bars representing standard deviation and symbols representing each of three replicates. (A) Doubling times of strains grown in liquid YEPD at 37 °C at 200 rpm. (B) Doubling times of strains grown in liquid YEP-Galactose at 37 °C at 200 rpm. (C) Statistical analysis of data from strains growing in YEPD. A one-way ANOVA was performed to determine if results were significantly different and a Dunnett's multiple comparisons test was performed to determine which means were significantly different from WT Wü284, asterisks represent degrees of significance. (D) Statistical analysis of strains growing in YEP-Galactose.

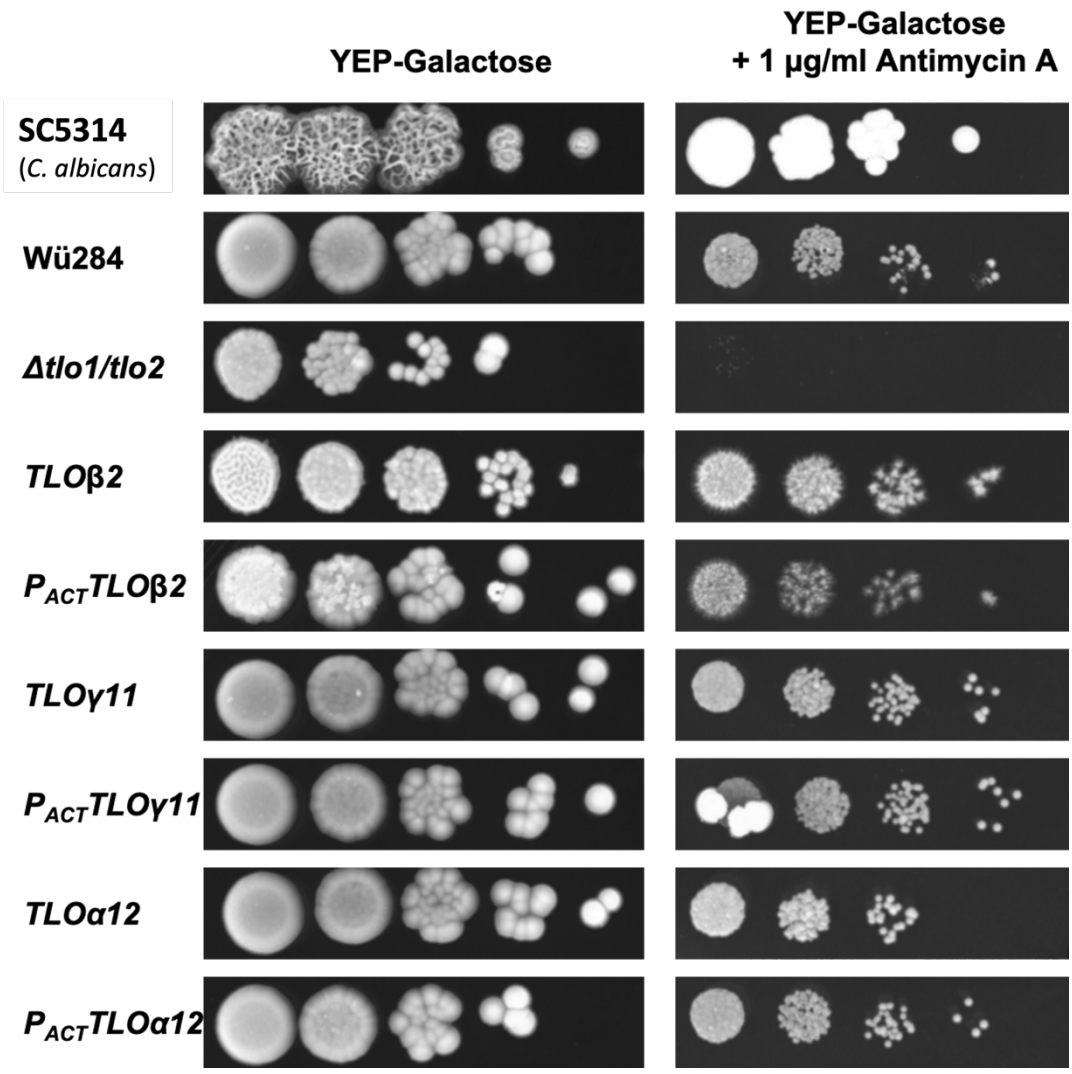


Figure 3.7 Growth of *C. dubliniensis* strains expressing *CaTLO*s on YEP-Galactose with Antimycin A

10-fold serial dilutions (10^6 – 10^2 cfu/ml) were plated in 5 µl volumes onto YEP-galactose agar with and without 1 µg/ml Antimycin A, an inhibitor of cellular respiration. Plates were incubated at 37 °C for 72h before photographing.

*CaTLO*s grew similarly to Wü284. The strain expressing *CaTLOγ11* under the *ACT1* promoter displayed a phenotype intermediate between the Wü284 strain and SC5314.

3.3.5 Biofilm formation

Biofilm formation in liquid Spider medium was measured at both 24 and 48 h using a crystal violet assay to measure biomass present, Figure 3.8. The two timepoints served to detect any effect the heterologously expressed *CaTLO*s had on biofilm formation or biofilm maintenance. Biofilm formation is generally increased by the addition of α clade *CaTLOα12*, either under the native or overexpressing *ACT1* promoter, however the natively expressed *CaTLOα12* strain was more similar to the *C. dubliniensis* WT at 48 h. *CaTLOβ2* strains are similar to the *C. dubliniensis* WT Wü284, except for P_{TLO1}*TLOβ2* displaying a lower level of biomass formed at the 24 h time point compared to the WT, and P_{ACT}*TLOβ2* showing less biomass than WT at 48h. For *CaTLOγ11* strains, the level of biofilm formation was significantly higher than that of WT at the 24 h time point. At 48 h *CaTLOγ11* under the native promoter showed a similar level of biomass present as the WT, but the P_{ACT1}*TLOγ11* strain formed significantly more biomass than WT.

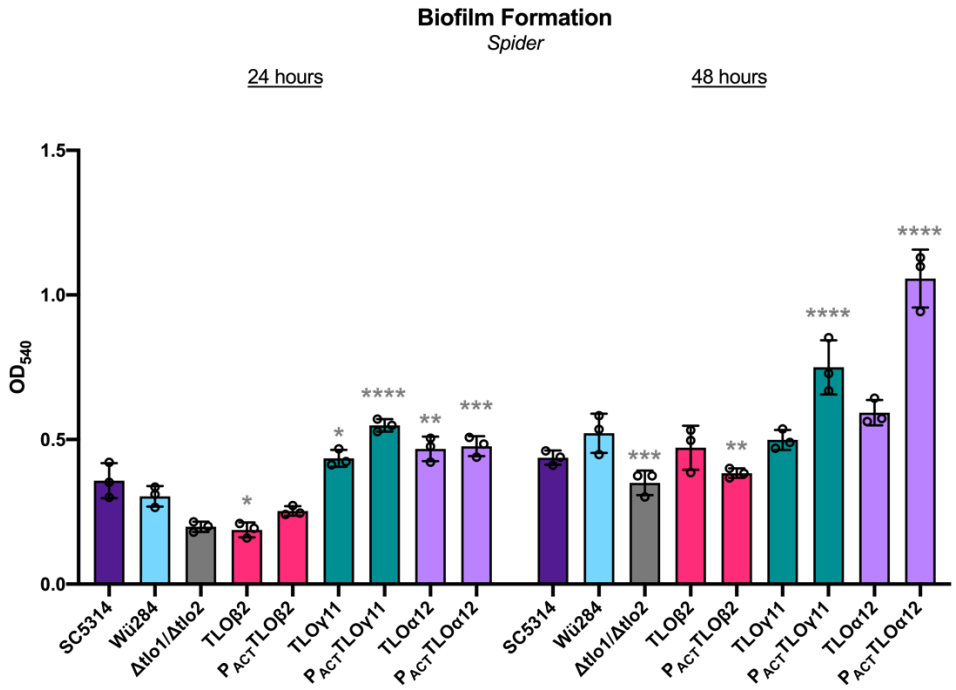
3.3.6 Cell wall perturbing compound spot plate assays

In order to determine if heterologous expression of *CaTLO*s in *C. dubliniensis* affected resistance to cell wall perturbing compounds, spot plate assays were performed using Calcofluor White and Congo Red at various concentrations. The most significant differences were observed with concentrations of 20μg/ml Calcofluor White, and 3 μg/ml Congo Red, Figure 3.9. Deletion of both native *CdTLO*s (*Δtlo1/Δtlo2*) is shown to reduce the ability of *C. dubliniensis* to grow in the presence of these agents. The *CaTLOβ2* expressing strains were both more susceptible to these compounds. Expression of *CaTLOγ11* and *CaTLOα12* had no significant effect on Calcofluor White resistance.

3.3.7 Oxidative stress minimum inhibitory concentration assays

Minimum inhibitory concentration assays in media containing decreasing levels of H_2O_2 or the organic peroxide, tBOOH, were carried out to determine the IC_{50} (concentration at which 50% of growth is inhibited) for each strain. Graphical representation of these results can be seen in Figure 3.10.

A



B

24 hours	SC5314 (*C. albicans*)	Wü284	Δtlo1/Δtlo2	TLOβ2	P_{ACT}TLOβ2	TLOγ11	P_{ACT}TLOγ11	TLOα12	P_{ACT}TLOα12
Mean (OD_{540})	0.3577	0.3038	0.1982	0.1873	0.2529	0.4348	0.5489	0.4676	0.477
S.D.	0.06058	0.0355	0.01787	0.02554	0.01626	0.02909	0.02165	0.04216	0.03457
Significantly different to Wü284	ns	-	ns	*	ns	*	****	**	***
P-value	0.7014	-	0.0715	0.0369	0.7552	0.0142	<0.0001	0.0013	0.0006

C

48 hours	SC5314 (*C. albicans*)	Wü284	Δtlo1/Δtlo2	TLOβ2	P_{ACT}TLOβ2	TLOγ11	P_{ACT}TLOγ11	TLOα12	P_{ACT}TLOα12
Mean (OD_{540})	0.4371	0.5219	0.3502	0.4715	0.3836	0.4987	0.7498	0.5925	1.057
S.D.	0.02472	0.06783	0.04226	0.07629	0.01631	0.03431	0.09341	0.04381	0.1
Significantly different to Wü284	ns	-	***	ns	**	ns	****	ns	****
P-value	0.2171	-	0.0007	0.7646	0.0086	0.9965	<0.0001	0.4031	<0.0001

.

Figure 3.8 Effect of expression of *CaTLOs* in *C. dubliniensis* on biofilm formation

The ability of strains to form biofilm on plastic surfaces (plastic 96 well plate) in liquid Spider medium was analysed. Biofilm formation was quantified by staining with crystal violet and measuring absorbance at 540 nm. This was done at 24- and 48-hour time points. (A) Graphical representation of biofilm formation data. Error bars represent standard deviation from the mean, and symbols represent values from three replicates. Asterisks denote statistically significant difference from WT Wu284. (B) Summary of data and statistical analysis (one-way ANOVA) of 24 h time point data. (C) Summary of data and statistical analysis (one-way ANOVA) of 48 h time point data.

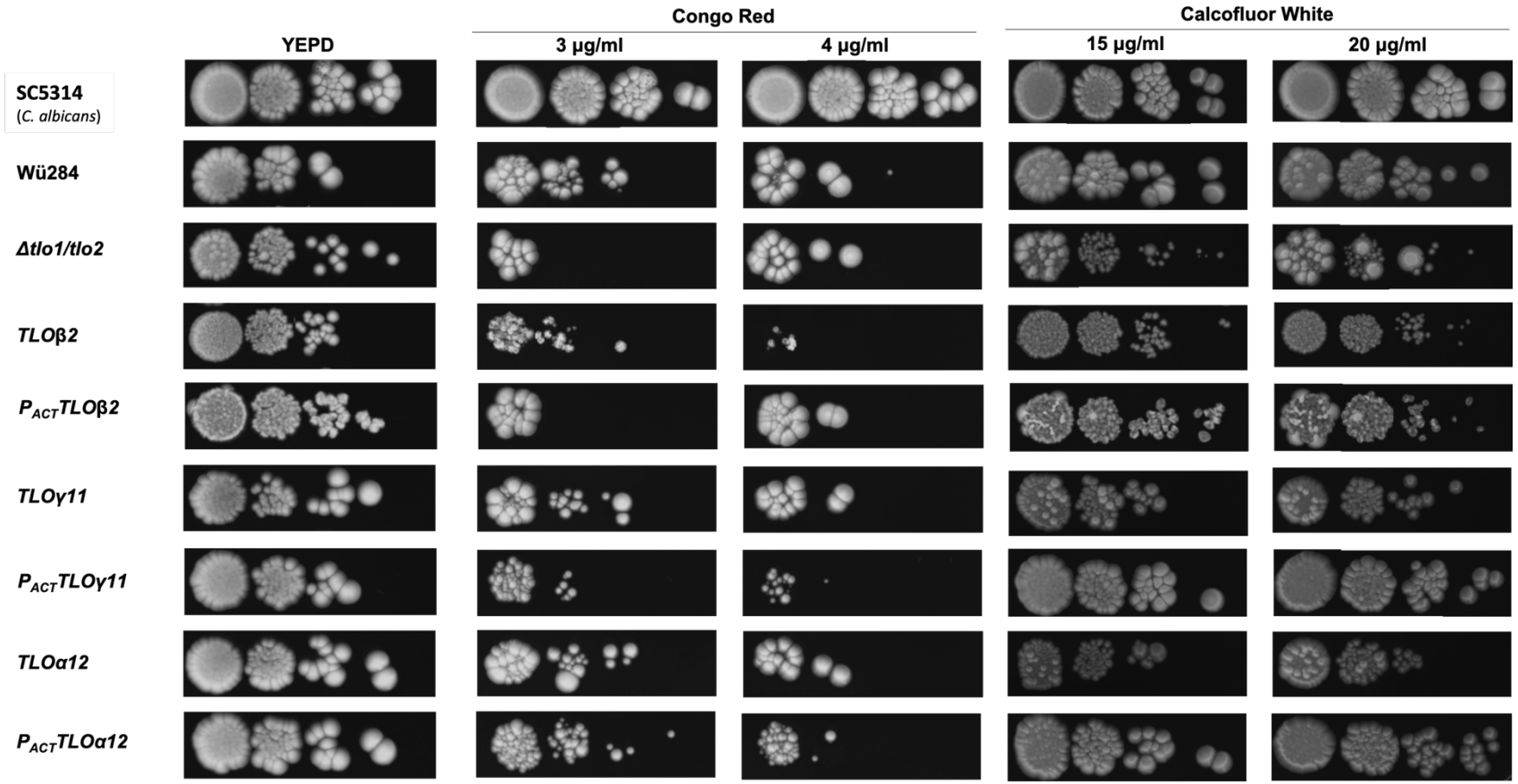
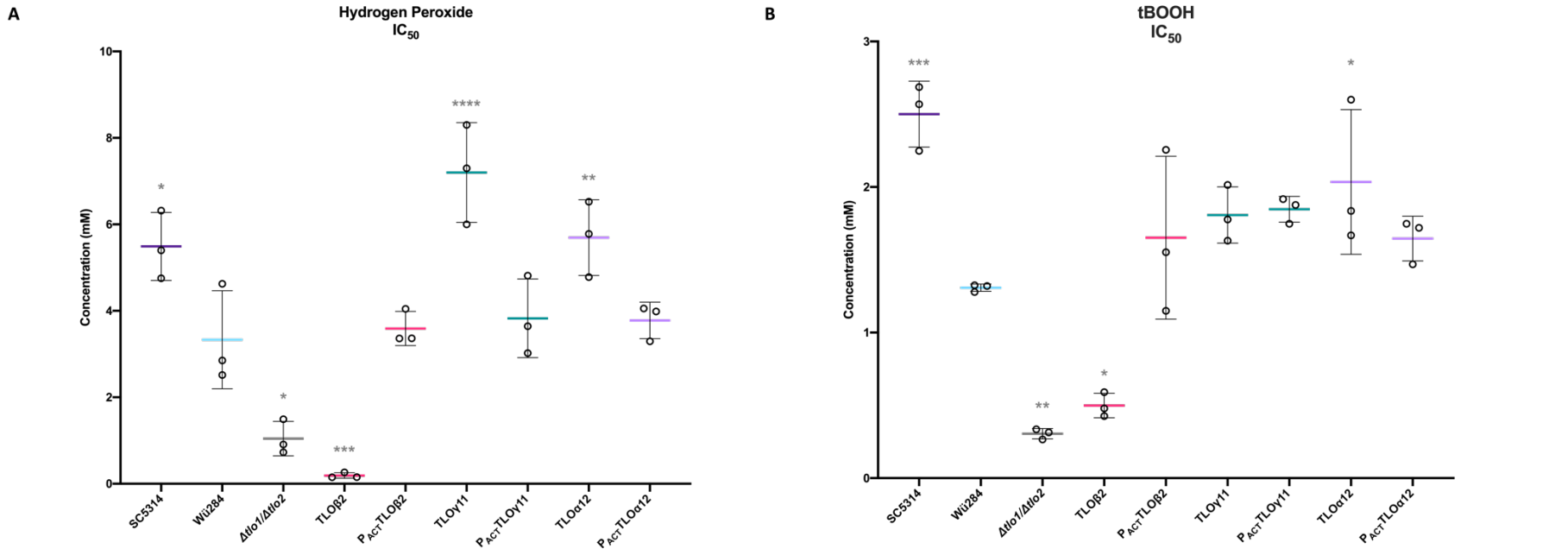


Figure 3.9 Growth of *C. dubliniensis* strains expressing *CaTLOs* on media containing cell wall perturbing compounds

10-fold serial dilutions (10^6 – 10^3 cfu/ml) were plated in 5 µl volumes onto YEPD agar containing different concentrations of either Congo Red or Calcofluor White to determine susceptibility of strains to these compounds. Plates were incubated at 37 °C for 72 h before photographing.



C

H_2O_2	SC5314 (*C. albicans*)	Wü284	Δtlo1/Δtlo2	TLOβ2	P_{ACT}TLOβ2	TLOγ11	P_{ACT}TLOγ11	TLOα12	P_{ACT}TLOα12
Mean (mM)	5.490	3.331	1.042	0.1899	3.590	7.199	3.826	5.694	3.778
S.D.	0.7875	1.134	0.3992	0.06370	0.3946	1.154	0.9092	0.8747	0.4205
Significantly different to Wü284	*	-	*	***	ns	****	ns	**	ns
P-value	0.0178	-	0.0115	0.0006	0.9979	<0.0001	0.9591	0.0089	0.9760

D

tBOOH	SC5314 (*C. albicans*)	Wü284	Δtlo1/Δtlo2	TLOβ2	P_{ACT}TLOβ2	TLOγ11	P_{ACT}TLOγ11	TLOα12	P_{ACT}TLOα12
Mean (mM)	2.501	1.308	0.3047	0.4986	1.652	1.807	1.847	2.035	1.645
S.D.	0.2266	0.02524	0.03549	0.08386	0.5597	0.1933	0.08888	0.4967	0.1533
Significantly different to Wü284	***	-	**	*	ns	ns	ns	*	ns
P-value	0.0004	-	0.0021	0.0131	0.5449	0.1951	0.1431	0.0282	0.5647

Figure 3.10 Susceptibility of *C. dubliniensis* strains expressing *CaTLOs* to oxidative stress

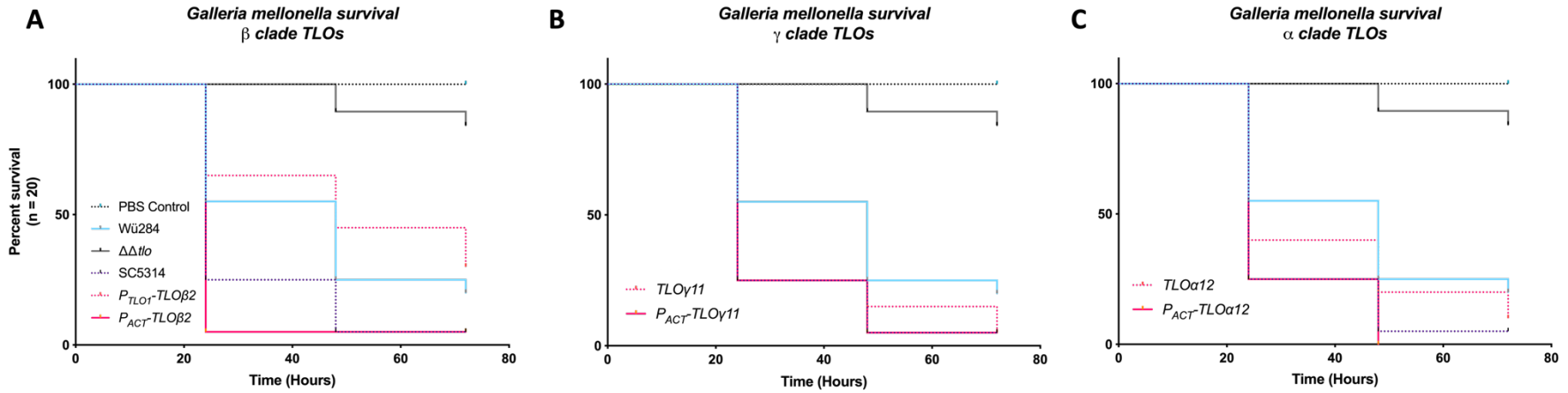
Minimum inhibitory concentration assays were performed by microtiter dilution to determine the IC_{50} of strains in oxidative stress inducing reagents. IC_{50} is the concentration at which growth is inhibited by 50%. (A) IC_{50} concentrations of each strain in hydrogen peroxide (H_2O_2), mean is indicated by the horizontal bar, with error bars representing standard deviation and symbols representing each of three replicates. Asterisks represent statistical significance in the difference in the IC_{50} of a strain from the IC_{50} of the WT Wu284 strain. (B) IC_{50} concentrations of strains in tBOOH. (C) Summary of data and statistical analysis (one-way ANOVA) of growth of strains in H_2O_2. (D) Summary of data and statistical analysis (one-way ANOVA) of growth of strains in tBOOH.

In H_2O_2, *C. dubliniensis* WT Wü284 was significantly more sensitive to oxidative stress than *C. albicans* WT SC5314. The *C. dubliniensis Δtlo1/Δtlo2* strain was more sensitive to oxidative stress induced by H_2O_2 than the WT *C. dubliniensis* strain. Expression of *CaTLOα12* and *CaTLOγ11* under the native promoter increased the IC_{50} statistically from that of the WT *C. dubliniensis* Wü284. This indicates that the presence of these *TLO*s reduces the susceptibility of *C. dubliniensis* to oxidative stress caused by H_2O_2. The expression of *CaTLOβ2* under the *CaTLOα1* promoter was found to decrease the IC_{50} to much lower than that of the WT, to even less than that of the *Δtlo1/Δtlo2 C. dubliniensis* strain. However, a similar result was not seen in the strain expressing *CaTLOβ2* under the *ACT1* promoter, where the IC_{50} was again similar to that of the WT. In tBOOH, the difference between the IC_{50} of the *C. albicans* and *C. dubliniensis* WT strains was significant with the *C. albicans* strain being much more resistant to the oxidative stress induced by tBOOH than *C. dubliniensis*. Deletion of both native *TLO* genes from *C. dubliniensis* (*Δtlo1/Δtlo2*) resulted in significantly increased susceptibility to tBOOH induced oxidative stress. Similarly to what was seen in the H_2O_2 MIC assay, the expression of *CaTLOβ2* under the *CaTLOα1* promoter increased susceptibility of the WT *C. dubliniensis* to oxidative stress, while the expression of *CaTLOβ2* under the high level *ACT1* promoter has no significant effect on the *C. dubliniensis* susceptibility level. Expression of *CaTLOγ11* under either a high or low level had some effect on increasing the resistance level, but this was not significant. The expression of *CaTLOα12* under the native promoter was found to be significantly more resistant to oxidative stress.

3.3.8 Galleria mellonella *infection model*

A *Galleria mellonella* wax moth larva infection model was used to determine if heterologous expression of *CaTLOs* could affect the virulence of the WT *C. dubliniensis* strain. Kaplan-Meier curves were generated in GraphPad Prism to show the survival of *G. mellonella* larvae after infection with the strains, Figure 3.11.

The WT *C. albicans* strain, SC5314, was found to be significantly more virulent than the *C. dubliniensis* WT Wü284. The *Δtlo1/Δtlo2* mutant strain was much less virulent than WT *C. dubliniensis.* Overall, it was seen that WT *C. dubliniensis* strains expressing any one of *CaTLOβ2*, *γ11* or *α12* under the *ACT1* promoter were statistically significantly more lethal than the WT *C. dubliniensis.* Expression of *CaTLO*s under their native promoters was not found to have a significant effect on the virulence of the WT *C. dubliniensis* Wü284 strain.



D

	SC5314 (*C. albicans*)	Wü284	Δtlo1/Δtlo2	TLOβ2	P_{ACT}TLOβ2	TLOγ11	P_{ACT}TLOγ11	TLOα12	P_{ACT}TLOα12
Significantly different to Wü284	*	-	****	ns	**	ns	*	ns	*
P-value	0.0462	-	<0.0001	0.3581	0.0066	0.0867	0.0462	0.3473	0.0114

Figure 3.11 Survival of *Galleria mellonella* infection model after infection with *C. dubliniensis* strains

Each *Galleria mellonella* worm was infected with 10^6 cells. Inoculation with the same volume of PBS was used as a negative control. In total 20 worms were infected with each strain/PBS. Results above are presented according to the results for each *CaTLO* clade. *C. dubliniensis* WT Wü284 and the *C. dubliniensis ΔΔtlo* strains were included for comparison. (A) Effect of addition of *CaTLOβ2* under either the *TLOα1* promoter or the ACT1 promoter. (B) Effect of addition of a CaTLO*γ11* under the control of either the native or *ACT1* promoter. (C) Effect of addition of a *CaTLOα12* under either the native of *ACT1* promoter. (D) Statistical analysis of results. Strains were compared to Wü284 using a Log-rank (Mantel-Cox) test. Degrees of significance are indicated by asterisks, and P-values are given.

3.4 Discussion

The aim of the experiments described in this chapter was to investigate what effect heterologous expression of selected representative *Candida albicans TLO*s (*CaTLOα12*, *CaTLOβ2* and *CaTLOγ11* would have on WT *C. dubliniensis* Wü284 and to determine if expansion of the *TLO* gene family in *C. dubliniensis* could increase fitness. In this chapter is assumed that the *CaTLO* mRNA expressed from the constructs is translated into protein, however, no formal quantification of protein production was performed. A summary of phenotypic analysis can be found in Figure 3.12 in the form of a heat map.

While all other *CaTLO* expressing strains of *C. dubliniensis* appeared morphologically identical to the Wü284 WT, the strains expressing *CaTLOβ2* were significantly different, Figure 3.3. These strains formed extremely wrinkled colonies on YEPD agar, and when examined under the microscope, they displayed a filamentous morphology. The low level P_{TLO1}*CaTLOβ2* appears to confer a slightly less wrinkled phenotype than the strain expressing at a higher level, P_{ACT1}*CaTLOβ2*. This difference can also be seen at the cellular level, where the P_{TLO1}*CaTLOβ2* strain was predominantly pseudohyphal, while the morphology of the P_{ACT1}*CaTLOβ2* expressing strain grew as is true hyphae. In general, in the *C. dubliniensis* strains described in this chapter, the increase in expression of *CaTLOβ2* the WT *C. dubliniensis* background appears to have a gene dosage-dependent effect, where the P_{ACT1}*CaTLOβ2* strain is more filamentous than the P_{TLO1}*CaTLOβ2*.

In the hyphal inducing conditions tested, the *CaTLOβ2* expressing strains were the only *CaTLO* supplemented *C. dubliniensis* strains to form hyphae. In 10% FCS these strains were constitutively hyphal, whereas SC5314 was observed to switch from the yeast form to the hyphal form, and then return to yeast form after around 4-6 h. This indicates that introduction of *CaTLOβ2* may interrupt the yeast-hyphal switching mechanism in *C. dubliniensis*. Strains constitutively expressing *CaTLOβ2* appear to be locked in the hyphal form.

Transcriptome profiling of a *C. albicans* infection model for invasive oral candidiasis by Zakikhany *et al.* in 2007 indicated that *CaTLOβ2* expression is significantly upregulated during epithelial infection (Zakikhany *et al.*, 2007). It is known that hyphae begin to form directly after epithelial contact during infection, and this could explain the upregulation of *CaTLOβ2* coinciding with the start of hyphal formation. In work

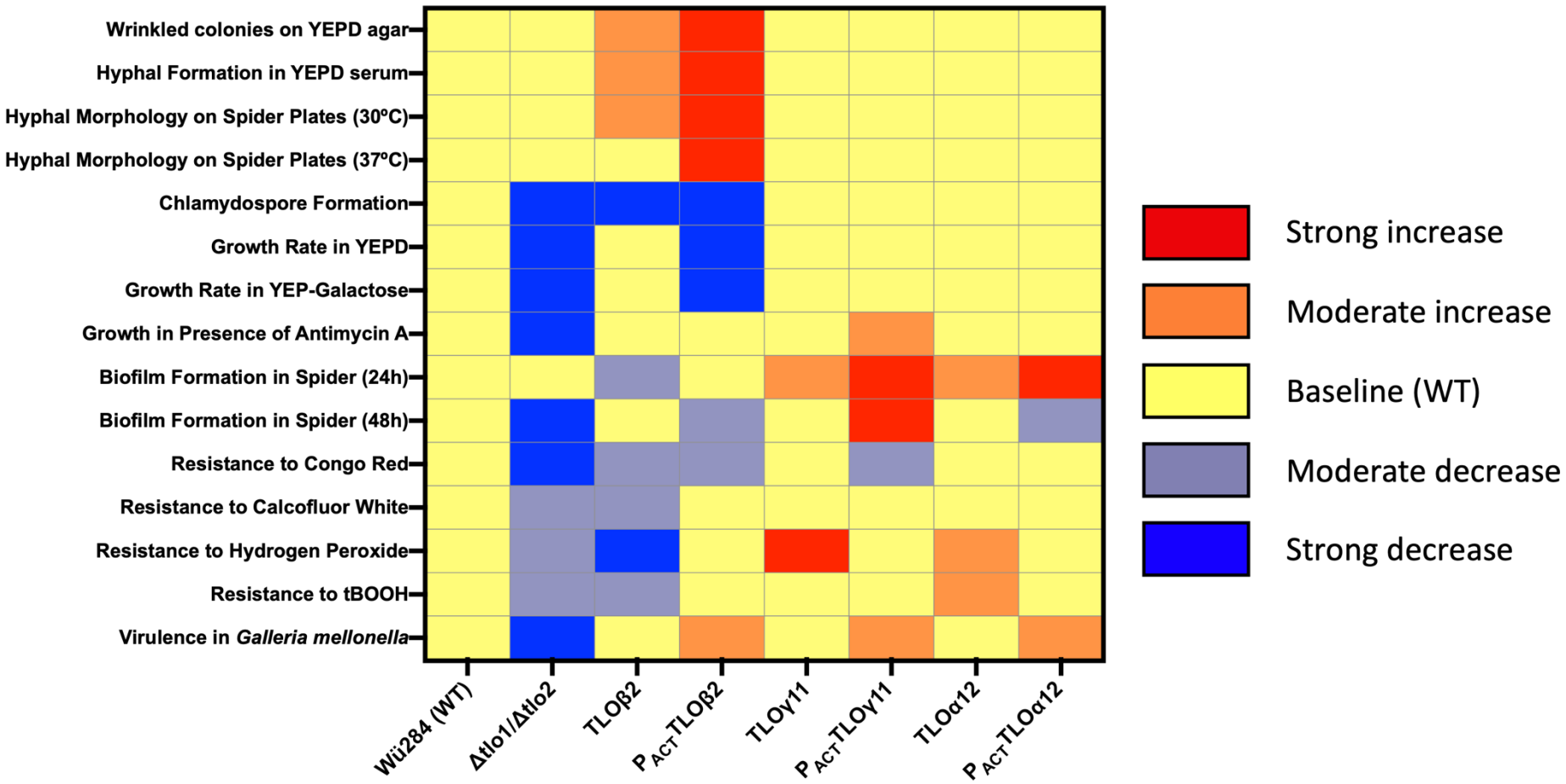


Figure 3.12 Heatmap of the effects of *CaTLO* expression in WT *C. dubliniensis*

Each phenotype was graded based on the observed difference from WT *C. dubliniensis* Wü284. These gradings were then plotted in Prism to form a heat map.

performed in 2018, Uppuluri *et al.* showed that a *Δtloβ2/Δtloβ2* strain of *C. albicans* SC5314 is defective for filamentation and biofilm formation. They demonstrated that this strain grew almost exclusively in the yeast morphology, even in hyphal inducing conditions. When this knockout was complemented with a single copy of the *CaTLOβ2* gene, the ability of the strain to form biofilm was somewhat restored, but not to the level of the wildtype strain (Uppuluri *et al.*, 2018). These data indicate that *CaTLOβ2* plays a significant role in the hyphal induction and the biofilm formation pathways in *C. albicans*. The biofilm formation pathway and the hyphal formation pathway are very closely linked, with many genes having roles in both pathways, and biofilms themselves being made up of a mixture of yeast and hyphal cells (Nobile and Johnson, 2015).

In gene knock down experiments performed by Dunn *et al.* (2018) the expression of *CaTLOβ2* in *C. albicans* WT SC5314 was reduced by placing one of the two native alleles of the gene under the control of a TET-ON promoter system. The knocked down strain did not show any defects in filamentation (Dunn *et al.*, 2018). This suggests that this mutant still had sufficient Tloβ2 for the normal control of filamentation.

One main difference between the hyphal development pathways in *C. albicans* and *C. dubliniensis* is that the latter requires a much harsher environment to induce filamentation. O'Connor *et al.* (2010) demonstrated that hyphal formation in *C. dubliniensis* requires nutrient starvation to induce the switch, whereas *C. albicans* can form hyphae much more readily, even in nutrient rich conditions (O'Connor *et al.*, 2010). It could be that the need for such a robust induction may be partly due to the reduced repertoire of *TLO* genes, or the lack of the specific *TLO* genes that are responsible for induction under these less stringent conditions.

During phenotypic analysis of the strains in this work, it appears that the filamentous *CaTLOβ2* expressing strains performed less well than other strains, sometimes even worse than the WT. This could be due to the expression of *CaTLOβ2* affecting these changes or may be due to the constitutive hyphal morphology of the cells. The hyphal cell wall composition of *Candida* is different to that of the yeast cell wall. Hyphal cell walls have a higher chitin content than the yeast cell wall. There is also less phosphodiesterified acid-labile β-1,2-linked manno-oligosaccharides in the hyphal form than the yeast form (Garcia-Rubio *et al.*, 2020). Congo Red is a chitin-binding dye that is thought to disrupt the links between chitin and β-glucan, thus disrupting cell wall integrity. Calcofluor White binds to the 1-4 β and 1-3 β polysaccharides bound to chitin

and is also proposed to disrupt cell wall integrity. In *S. cerevisiae*, mutants with less chitin in their walls are typically more resistant to Calcofluor White and Congo Red, and those with more chitin in their cell walls are generally more sensitive. It was seen however that not all *S. cerevisiae* mutants with larger amounts of chitin in their walls were sensitive to Calcofluor White. This indicates that there could be another mechanism of antifungal action related to this compound (Ram and Klis, 2006). The *CaTLOβ2* strains showed increased susceptibility to Congo Red and Calcofluor White in spot plate assays, Figure 3.9, possibly due to the increased chitin levels in hyphal cell walls. The hyphal composition of the cell wall of pseudohyphae is unclear, the fact that it is not known if this morphology is a terminal morphology or a transitional state between yeast and hyphae (Noble *et al.*, 2017), or in the case of genetic mutants, a defect in cell division, means there could possibly be many different pseudohyphal strains with differing cell wall make ups. While work by Chattaway *et al.* in 1968 found that there was a higher content of chitin in the mycelial form of *C. albicans,* i.e. the hyphal cell wall, a study in 1959 by Kessler and Nickerson found that there was no difference in cell wall make up in a morphologically yeast-like strain and a pseudohyphal strain, which they called a 'non-septate pseudo mycelium' (Chattaway *et al.*, 1968, Kessler and Nickerson, 1959). While these studies are quite dated, it does indicate that pseudohyphal cell walls are more likely to be more similar in composition to yeast cell walls than hyphal cell walls.

The ability to form chlamydospores was impeded in the *CaTLOβ2* strains, as well as the *Δtlo1/Δtlo2* strain, Figure 3.5. It could be that the *CaTLOβ2* strains are stuck in the hyphal morphology and are unable to activate the chlamydospore formation pathway. Many genes involved in the hyphal formation pathways in *Candida* are also involved in the chlamydospore formation pathway such as *EFG1*, which plays a major role in the yeast to hyphal transition. *EFG1* mutants are unable to form hyphae in inducing conditions and are also unable to form chlamydospores (Sonneborn *et al.*, 1999). It could be that the *CaTLOβ2* is acting on these pathways and masking or blocking any effects of chlamydospore induction, and therefore the hyphal morphology persists.

It may be the case that the hyphal morphology of the *CaTLOβ2* expressing strains is causing slower growth rates than the WT, Figure 3.6. An *Δnrg1/Δnrg1* mutant strain of *Candida albicans,* which grows as wrinkled colonies on YEPD with a filamentous cellular morphology, was much slower growing than the WT SC5314 parent (Braun *et al.*, 2001). However, another filamentous mutant of *C. albicans*, *Δtup1/Δtup1*, does not display a

reduced growth rate in YEPD at 37 °C (Braun and Johnson, 1997). This indicates that while disruptions in the hyphal development pathways can cause reduced growth rate, it is not necessarily the filamentous morphology itself that is causing the slow growth.

The phenotypes affected by *CaTLO* expression in the *C. dubliniensis* WT background did not reveal any clade specific patterns. Except for the effect of expressing *CaTLOβ2* on hyphal morphology, *CaTLO* genes from each clade affected a wide range of phenotypes. Expression of *CaTLOγ11* and *CaTLOα12* in the *C. dubliniensis* WT background conferred some advantages to the strains in various challenges, but there was no clear clade specific, or expression level specific trend observed. This indicates that there may be a level of functional redundancy between *CaTLO* genes. This is likely a result of the evolution of all the *CaTLO* genes from a common ancestral locus (Anderson *et al.*, 2012).

Dunn *et al.* (2018), showed that knock down of *CaTLOγ11* or *CaTLOα12* results in a slower growth rate in YEPD at 30 °C. They performed varied phenotypic analysis on their knockdown strain set and constructed a phenotypic relatedness chart between the strains. This generated two main clusters of *TLO*s that contained genes from varying clades. Principal component analysis of the data however yielded more clade ordered clustering, and the *α* clade *TLOs* formed a cluster, while the *γ* clade *TLO*s formed two distinct clusters. *CaTLOβ2* is the only member of the *β* clade, and this sat apart from the other clusters. Dunn *et al.* use these data to speculate that *CaTLO* genes can have clade specific functions but can also have gene specific functions (Dunn *et al.*, 2018).

An expansion on the work described in this chapter could involve creating a strain set of *C. dubliniensis* mutants with each of the *CaTLO* genes expressed in the WT. Performing phenotypic analysis and then clustering the strains based on the data could highlight trends in clade specific responses, similar to what was performed by Dunn *et al.* above.

It does appear that the expression level of the *CaTLO* constructs in WT *C. dubliniensis* plays a role in virulence in the *Galleria mellonella* virulence assay, Figure 3.11. Here, only the strains expressing *CaTLO* genes under the *ACT1* promoter displayed an increase in virulence compared to the WT *C. dubliniensis* strain. The fact that WT *C. albicans* SC5314 has 14 *TLO* genes, suggests that it is likely that there is much more Tlo protein present in the *C. albicans* cell compared to that in the *C. dubliniensis* cell. Estimates suggest that there is up to 10 times the amount of free Tlo protein in the *C. albicans* cell compared to that which is Mediator associated, while *C. dubliniensis* does not have any free Tlo. Gene

dosage of *TLO*, regardless of clade, could be partly responsible for the increased virulence of *C. albicans* compared to *C. dubliniensis*. However, as mentioned previously, quantification of the protein level expressed from the *CaTLO* constructs was not performed. Quantification could be performed incorporating a protein tag to the constructs to allow for detection and quantification of expressed proteins by Western Blot.

It has been shown that expression levels of Tlo proteins in *C. dubliniensis* are tightly regulated. A free pool of CdTlo1 protein cannot be generated via overexpression – as demonstrated by Haran *et al.* in 2014. This study however demonstrated that a free pool of CdTlo2 could be generated via overexpression from native loci (Haran *et al.*, 2014). In this current study we assume that the genes heterologously expressed from the *ACT1* promoter are over expressed. However, we have no way of knowing if there is excess Tlo protein that is not incorporated into the Mediator complex. In other words, are the phenotypes we observed due to the *C. albicans* Tlo protein replacing *C. dubliniensis* Tlos in Mediator or are free Tlo proteins acting independently of Mediator? These questions could be answered in future work by repeating these experiments with tagged Tlo proteins. This would permit protein expression levels to be determined and mass spectrometry experiments to be performed to identify which other proteins the Tlos are interacting with. Identifying the role of Mediator-free Tlos could also be investigated using a *Δmed3* mutant background, which would prevent the Tlos from being incorporated into Mediator.

The experiments detailed in this chapter suggest that *TLO* genes play a role in a wide array of functions in the *Candida* cell. If *CaTLO* genes were introduced to another *Candida* species, would there be similar effects? Though *C. albicans* and *C. dubliniensis* are the only two *Candida* species to form true hyphae, what would be the result of expressing *CaTLOβ2* in *C. parapsilosis*, *C. tropicalis* or *C. glabrata,* all of which only have a single *TLO* gene?

Further work in these strains is needed to answer more detailed questions about the role that the CaTlo proteins are playing in the *C. dubliniensis* cell. Do the CaTlos outcompete the native CdTlos for space in the Mediator complex? The CaTlos could interact with different factors in *C. dubliniensis* than the native CdTlos do to enact their function.

The work detailed in this Chapter confirms that the expansion of the *TLO* family of a *Candida* species can dramatically influence phenotype and virulence. There is evidence for functional diversity between the clade of *TLO* in *C. albicans,* as well as some functional overlap. These data also indicate that gene dosage is an important factor with regards to *TLO* genes. Further investigation may explain the link between the highly expanded *TLO* family in *C. albicans* and increased virulence. To further investigate the role *TLOs* in *Candida* biology it was decided to first attempt to delete the complete *TLO* repertoire in *C. albicans*. This would allow observation of the phenotype of a mutant without Tlo protein, as well as providing a clean background to reintroduce individual *TLO* genes to investigate their functions.

Chapter 4

Deletion of the entire *TLO* gene family from *Candida albicans* using CRISPR-Cas9 Mutagenesis

4.1 Introduction

4.1.1 Challenges in deleting the entire Candida albicans TLO *family*

The fact that *C. albicans* contains many more *TLO* genes than other *Candida* species presents a challenge. The lab strain SC5314 possesses 14 *TLO* genes (Jackson *et al.*, 2009) (with two copies of each due to its diploid genome), meaning that using traditional molecular techniques, up to 28 separate transformation/recombination events would be required to delete both copies of all 14 genes. Traditional gene deletion techniques would also require that resistance markers be recycled, with screening and confirmation of deletions being performed after each round of transformation. Clearly such an undertaking would have been technically very challenging, however, fortunately, the development of the CRISPR-Cas9 mutagenesis method offered a possible solution to this conundrum.

4.1.2 CRISPR-Cas Systems

The CRISPR-Cas system is an adaptive immune system in bacteria and archaea. The acronym CRISPR stands for Clustered Regularly Interspaced Short Palindromic Repeats, describing a series of direct repeats, about 25-35 bp in length, separated by short unique sequences of about 30-40 bp that make up a CRISPR locus (Koonin and Makarova, 2019). When foreign DNA (e.g. an infecting bacteriophage) is detected, the host can take a short segment of that DNA and integrate it into a CRISPR locus, in between the repeats. The CRISPR locus is often flanked by accessory genes, which encode proteins with roles in this immune-like response. The main family of genes are the *cas* genes (CRISPR associated), but others can be involved also (Koonin and Makarova, 2019). When foreign DNA matching the stored sequence is encountered again, CRISPR derived RNA (crRNA), in association with the Cas proteins, can detect and destroy the foreign DNA, thus providing immunity to bacteriophage containing those DNA sequences (Wiedenheft *et al.*, 2012). This system of proteins was first described in 1987, but in recent years they have been harnessed to facilitate precision editing and manipulation of genes, including the deletion of large gene families (Hsu *et al.*, 2014). There are many different types of CRISPR systems, divided mainly depending on the Cas proteins they use. These are discussed in more detail below.

While CRISPR may seem like a new topic in biology, the first mention of the CRISPR locus was in a 1987 paper from a group in Japan studying the *iap* gene in *E. coli*. The authors noted a section of the genome downstream from their gene of interest with 29 bp repeats interspaced with 32 bp unique sequences (Ishino *et al.*, 1987), what we now know to be the CRISPR locus. However, it wasn't until 2002 when the term CRISPR was coined. In-depth analysis by molecular biologists, in particular by the Nobel Laureates Jennifer Doudna and Emmanuelle Charpentier and their teams, has led to advances in understanding these bacterial and archaeal immune systems, leading us to where we are today, with the ability to edit the genomes of many cell types with these systems with relative ease (Hsu *et al.*, 2014). A timeline can be found in Figure 4.1 illustrating the history of CRISPR.

CRISPR-based precision mutagenesis technology has heralded a new era in molecular biology and has been applied to a wide range of medical and biological questions. One area in which CRISPR has been utilised in the clinical setting has been the application of gene editing in the development of novel cancer immunotherapies. Adoptive T cell therapy involves infusing genetically engineered T cells into a patient, and this can increase a patient's natural anti-tumour response (Stadtmauer *et al.*, 2020). The CRISPR engineered T cells have been administered to patients with refractory cancer. This trial has suggested that CRISPR edited T cells can persist for up to nine months, and that CRISPR-Cas9 is a safe method to use for editing human T cells.

Human genome editing doesn't come without controversy. In 2018, He Jiankui announced to the world that he had 'successfully' used CRISPR-Cas9 gene editing to modify the genomes of two babies in China, Lulu and Nana. This experiment was performed on IVF embryos which were later implanted into the mother. He's aim was to generate genetic immunity to HIV, via mutation of the *CCR5* gene in the embryos (Cohen, 2019). *CCR5* encodes a cell surface protein on lymphocytes and other cell types, that acts as a chemokine receptor for signalling and immune response modulation, however its role isn't fully understood yet. It is also a HIV coreceptor that allows the virus to enter the cells (Huttenrach *et al.*, 2005).

There is evidence that people who naturally carry a 32 base pair deletion in the *CCR5* gene have immunity to infection by HIV, however, there is also evidence for the deletion in *CCR5* predisposing carriers to encephalitis and death. There is no information as of

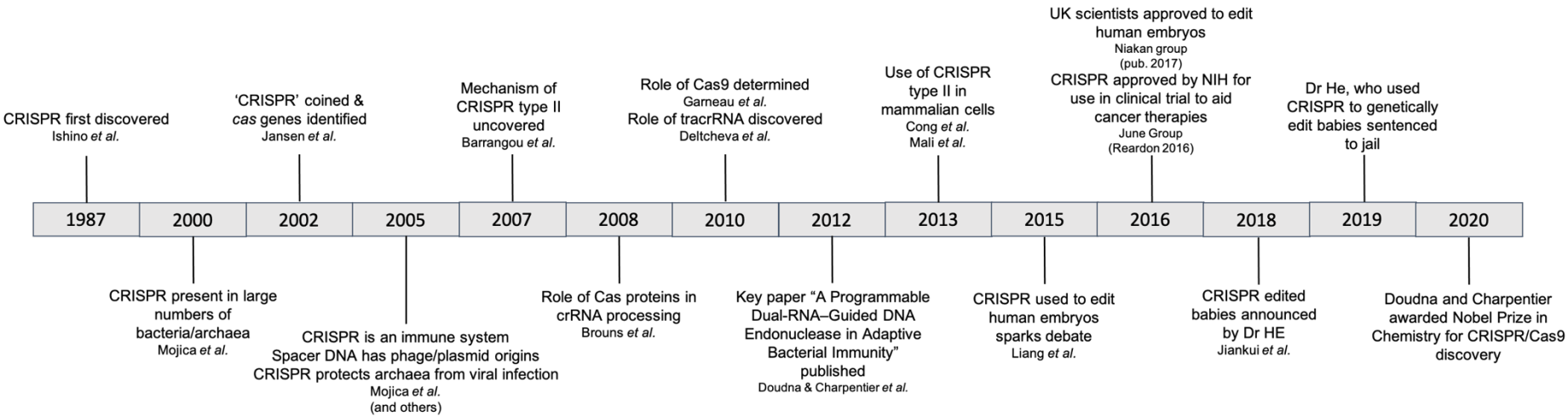


Figure 4.1 Timeline of discoveries relating to CRISPR

This timeline highlights some of the major discoveries that broadened our understanding of CRISPR systems. From the initial discovery of clustered, interspaced repeats in *E. coli* in 1987, up to the awarding of the 2020 Nobel Prize for CRISPR gene editing. This figure was compiled based on information in (Koonin and Makarova, 2019) as well the original publications (Ishino *et al.*, 1987, Normile, 2019, Cohen, 2019, Ledford and Callaway, 2020, Jinek *et al.*, 2012, Mojica *et al.*, 2000, Jansen *et al.*, 2002, Mojica *et al.*, 2005, Barrangou *et al.*, 2007, Brouns *et al.*, 2008, Garneau *et al.*, 2010, Deltcheva *et al.*, 2011, Cong *et al.*, 2013, Mali *et al.*, 2013, Liang *et al.*, 2015, Fogarty *et al.*, 2017, Reardon, 2016).

yet of any possible off target effects of He's mutagenesis experiment, or if there have been any side effects related to the deletion in *CCR5* itself (Cohen, 2019). In 2019, He was sentenced to 3 years imprisonment for illegal medical practices, He was also found to have falsified documents and misled other doctors (Normile, 2019). He's experiments generated outrage around the world, but also opened the door for many conversations around the bioethics of editing human genomes and embryos.

CRISPR technology has also been deployed as a diagnostic tool for a variety of microbial and viral infections. The ability of the system to identify the presence of a small segment of DNA or RNA with high specificity has proven very useful for diagnosing infection by certain microorganisms. Sherlock Bioscience is a biotechnology company in Boston, Massachusetts, whose flagship SHERLOCK technology uses CRISPR coupled with a Cas12 or Cas13 protein (for DNA and RNA targeting respectively) to detect genetic sequences from any organism with high specificity. SHERLOCK stands for Specific High-Sensitivity Enzymatic Reporter UnLOCKing and involves the detection of target genetic sequences which triggers the cleavage of a reporter sequence that generates a detectable signal, typically fluorescence or a signal detectable on a lateral flow device (Kellner *et al.*, 2019). During the Covid-19 pandemic, such diagnostics have proven essential in the fight against the virus, where quick and reliable detection of infections is key to preventing spread.

As referred to previously, in 2020, the Nobel Prize in Chemistry was awarded to Charpentier and Doudna for 'the development of a method for genome editing' (Ledford and Callaway, 2020). Eight years since the publication of their first paper on CRISPR, and after countless uses of their technology around the world, CRISPR was heralded as having taken life sciences into the future and being one of the greatest discoveries of our time.

4.1.3 How CRISPR-Cas systems work

The CRISPR system of adaptive immunity is split into three parts, (i) the adaption stage (acquisition and storage of unique sequences), (ii) the expression and processing stage (transcription of the CRISPR locus and processing of the RNA into crRNA) and (iii) the interference stage (the RNA directs destruction of target DNA by an effector complex) (Jinek *et al.*, 2012).

In the adaption stage, a complex made up of Cas proteins binds foreign DNA (e.g. bacteriophages) invading the bacterial cell and induces two sets of double stranded breaks in the DNA. Cleavage is induced after the Cas proteins recognize a 2-4 nucleotide motif called the protospacer adjacent motif (PAM). This releases a small segment of DNA, which is termed the protospacer, which is then integrated between two of the repeat sequences in the CRISPR locus. The invading DNA in the CRISPR locus is now called the spacer (Koonin and Makarova, 2019).

The CRISPR locus is transcribed as a single transcript, termed pre-crRNA, and this must be processed into mature crRNA. These steps make up the expression and processing stage. Depending on the CRISPR system, a complex of Cas proteins may be used for processing, a single dedicated Cas protein, or the hosts own RNase enzymes may be recruited for the processing stage (Koonin and Makarova, 2019). The different types and classes are described in Section 4.1.4 and shown in Figure 4.2.

The interference stage is the targeting and destruction stage. The crRNA is used as a guide (gRNA) and recognises the protospacer sequence in the invading DNA (assuming that the DNA sequence had previously been encountered). Cleavage requires complementarity between the gRNA and the protospacer, as well as the presence of the PAM. The target is then cleaved by Cas proteins. These may be a part of a larger effector complex used in the processing stage, or it may be a separate Cas protein (Jinek *et al.*, 2012, Koonin and Makarova, 2019).

4.1.4 Classes and types of CRISPR-Cas systems

The CRISPR systems found in bacteria and archaea fall into two main classes. Systems that use multi-subunit effector complexes fall into class one, while those that use a single Cas protein with multiple and different active domains in the interference stage fall into class two. The systems can then be subclassified into types based on more subtle difference in effector proteins. Type I, III and IV systems fall into class one, and type II, V and VI are in class two, see Figure 4.2.

Another difference between class one and class two CRISPR systems is in the processing stage. In class 1, the Cas6 protein is responsible for processing the pre-crRNA, whereas in class 2 this is not the case. In type II systems, the host cell RNase III (separate to the CRISPR system) is recruited to in the processing stage where it works along with the

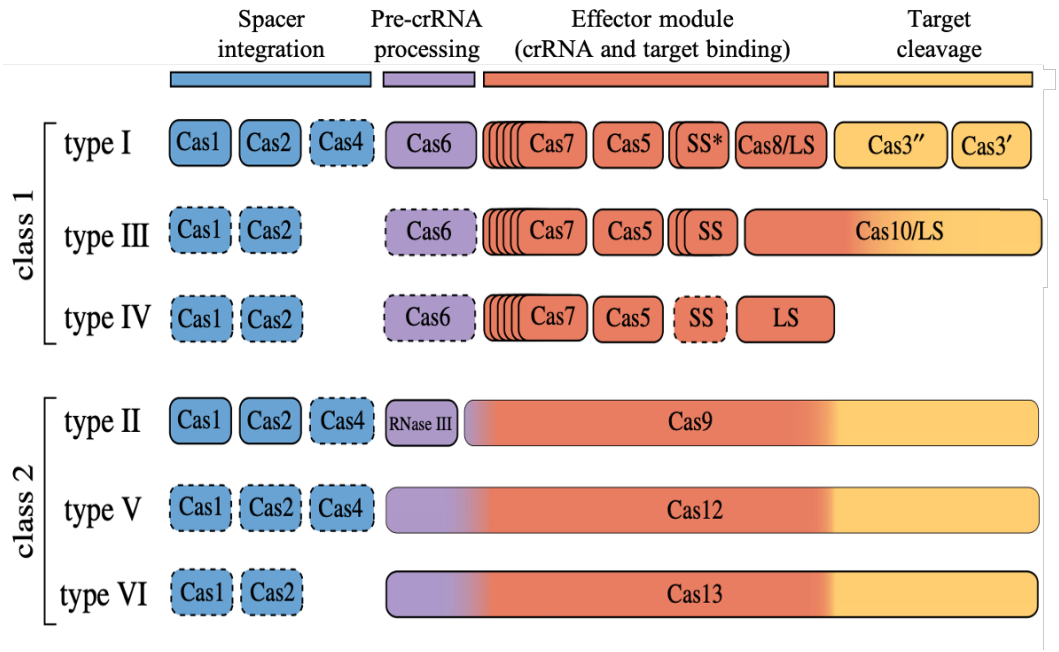


Figure 4.2 CRISPR system classifications

Here the two classes of CRISPR systems are shown with the specific types of system that fall into each class. Class 1 contains systems that use multiple proteins in an effector complex for crRNA, target binding and target cleavage, while class 2 systems use a single protein with multiple domains for these processes. Further classification into types is based on protein make-up of the systems. This figure shows the proteins that are involved at each stage of the CRISPR system for each type of system, with colour coding indicating which stage of the process they are involved in. This figure was adapted from a figure in (Koonin and Makarova, 2019).

Cas9 protein to process the pre-crRNA (Jinek *et al.*, 2012). The type II CRISPR system is the system used in this work. A more detailed description of its method of action can be found in Section 4.1.5. In type V and VI CRISPR systems, the processing step is carried out by the same Cas protein as the interference stage, Cas12 and Cas13 respectively (Koonin and Makarova, 2019).

4.1.5 CRISPR-Cas9 and gene editing

Type II CRIPSR systems use a single Cas protein (Cas9) in the processing and interference stages. The system also involves the RNase III ribonuclease in the processing step. Cas9 is a large protein that contains two different endonuclease domains, one with a RuvC-like domain, and an HNH domain (Wiedenheft *et al.*, 2012), with each domain cutting a single strand of DNA each. The HNH domain cuts the strand that is complementary to the crRNA, while the RuvC-like domain cuts the non-complementary strand (Jinek *et al.*, 2012). The type II CRISPR system also relies on another component, a small piece of non-coding, trans-activating RNA (tracrRNA). The tracrRNA is complementary to the crRNA and is involved in triggering the processing of pre-crRNA by RNase III in the processing stage and is also needed to activate the DNA cleavage of Cas9. However, it has been shown that an engineered piece of RNA that has features of both the crRNA and the tracrRNA can be used to direct the Cas9 enzyme and activate double strand cutting (Jinek *et al.*, 2012).

A basic schematic of the operation of the CRISPR-Cas9 system for genetic manipulation in relation to the experiments carried out in this work can be found in Figure 4.3. Details of the system used for gene editing in this work are described in detail in the following section.

4.1.6 Gene editing in Candida albicans *via a specialised CRISPR-Cas9 system*

The optimisation of CRISPR-Cas9 for genome editing in *Candida albicans* was not simple, mainly due to codon usage in CTG clade. The Fink group was able to solve these problems through synthesising a version of the *CAS9* gene that did not use the CTG codon, thus preventing the misincorporation of serine into the amino acid chain (Vyas *et al.*, 2015, Uthayakumar *et al.*, 2020). The fact that *C. albicans* does not have expression systems for small RNAs meant that the gRNA had to be placed under the control of the

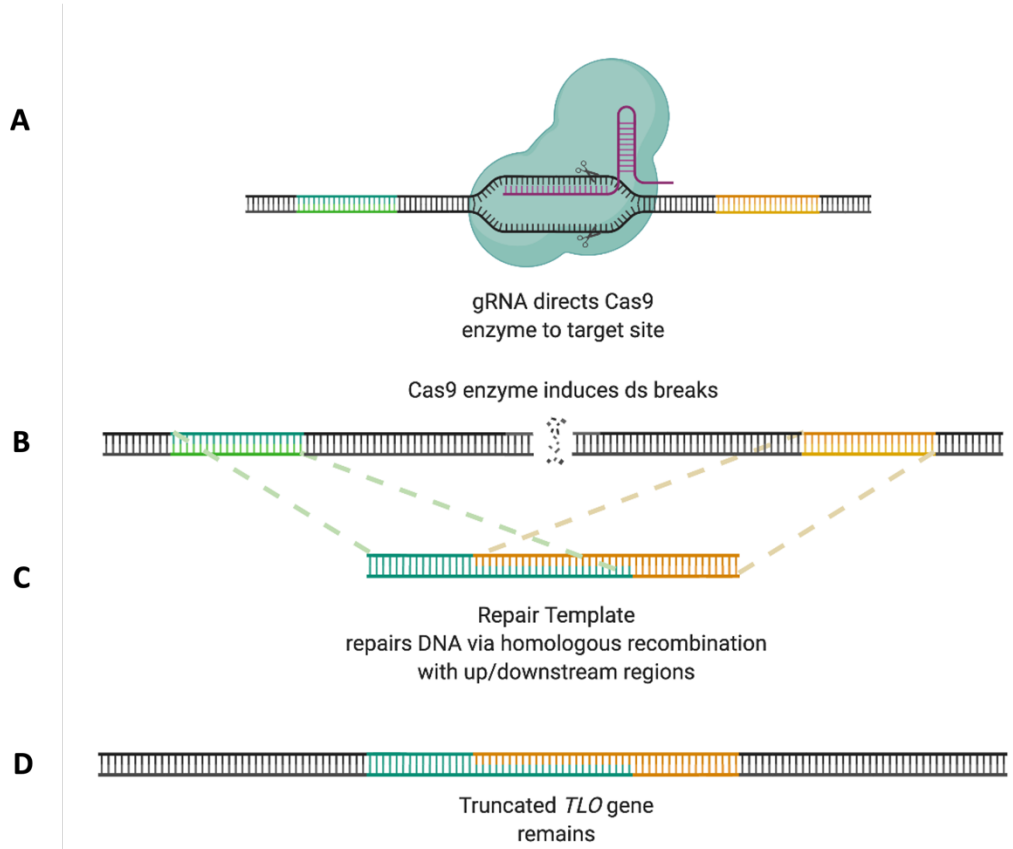


Figure 4.3 CRISPR-Cas9 system for gene editing

This is a simple outline of how the type II CRISPR-Cas9 system can be used in genetic engineering. (A) The gRNA (purple) contains elements of the crRNA and tracrRNA so that only a single RNA guide is needed. A unique 20 nucleotide sequence within the gRNA couples with and directs the Cas9 enzyme to a target site in the genome (*MED2* domain of the *TLO* gene family here, black) and induces double stranded breaks (B). A repair template is supplied ((C) green and orange) with homology to up/downstream regions of the break site (green and orange), and so homology directed repair can occur. It is possible to use the repair template to introduce a desired sequence at this point. If no repair template is supplied, non-homologous end joining occurs. In this work, the repair template is used to create a truncated TLO gene (D). This figure was created in BioRender based on information in (Koonin and Makarova, 2019, Nguyen *et al.*, 2017).

SNR52 RNA Polymerase III promoter (Vyas *et al.*, 2015). The first successful use of CRISPR-Cas9 in *C. albicans* was performed by the Fink group, where they successfully knocked out both copies the *ADE2* gene, giving a red phenotype to cells when both copies are inactive (Vyas *et al.*, 2015).

Current advances in using CRISPR in *C. albicans* include using CRISPR interference (CRISPRi) and CRISPR activation (CRIPSRa) to alter the transcriptome without directly editing the genome itself (Uthayakumar *et al.*, 2020). This approach uses a nuclease-dead Cas9 that has the ability to target DNA based on a gRNA sequence, but does not have the ability to cleave. This nuclease-dead Cas9 can then be fused to transcriptional activators (e.g. VP64), or repressors (e.g. Nrg1) in order to control transcription of genes targeted by the guide (Uthayakumar *et al.*, 2020, Roman *et al.*, 2019).

A specialised system for genetic engineering in *C. albicans* was used in this work. This system developed by Nguyen *et al.* (Nguyen *et al.*, 2017) has been shown to successfully generate homozygous transformants in *C. albicans* using a single transformation. Benefits of this system include the ability to recycle the CRISPR system and antibiotic resistance marker, as well as the ease with which the custom gRNA can be implemented in the system.

The system works by introducing a gRNA cassette and a Cas9 cassette into the cell by transformation, where homologous recombination joins the two linear cassette fragments and forms the working system. Each fragment contains one half of the nourseothricin resistance gene (NAT cassette) which is the basis for the recombination event. Each fragment also contains one half of the *LEU2* gene, which allows for integration of the system into the *LEU2* locus of *C. albicans*. The ability to recycle the CRISPR system and the NAT cassette results from transforming the system into a *LEU2/Δleu2* heterozygous mutant strain. With resultant transformants being NAT^{+}/LEU^{-}, growing transformants on leucine negative media drives the reconstitution of the *LEU2* gene, resulting in the excision of the CRISPR cassette and NAT marker, ultimately resulting in the formation of a NAT^{-}/LEU^{+} strain, in what the creators call the LEUpOUT method (Nguyen *et al.*, 2017).

Plasmids containing the components of the CRISPR-Cas9 system used in this study were supplied by Aaron Hernday's group (Nguyen *et al.*, 2017) and a schematic diagram of the process is shown in Figure 4.4. The plasmids can be amplified, and the appropriate

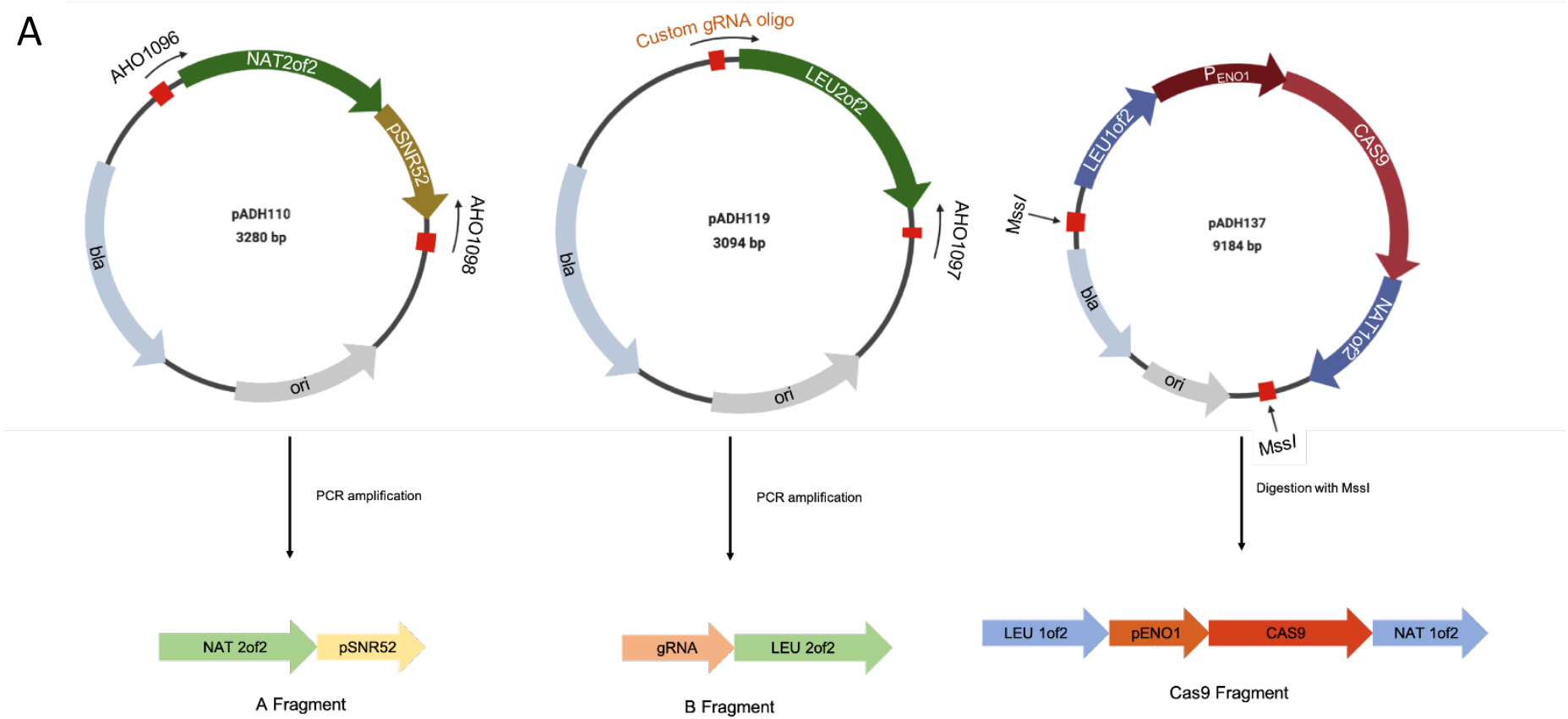


Figure 4.4 Specialised *C. albicans* CRISPR-Cas9 system (A)

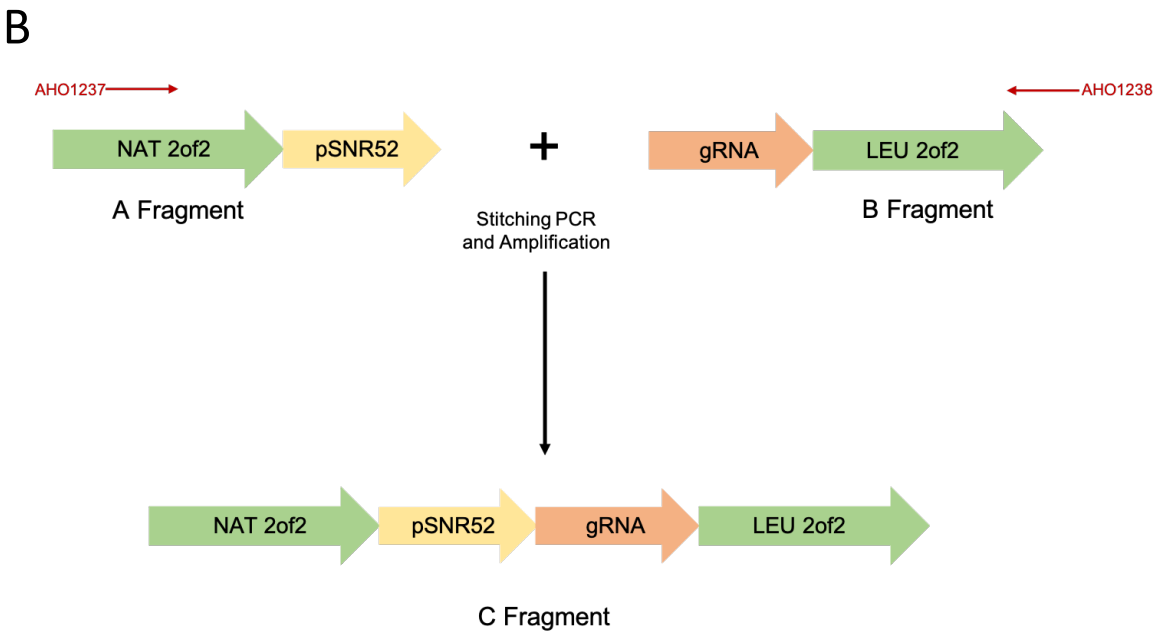


Figure 4.4 Specialised *C. albicans* CRISPR-Cas9 system (B)

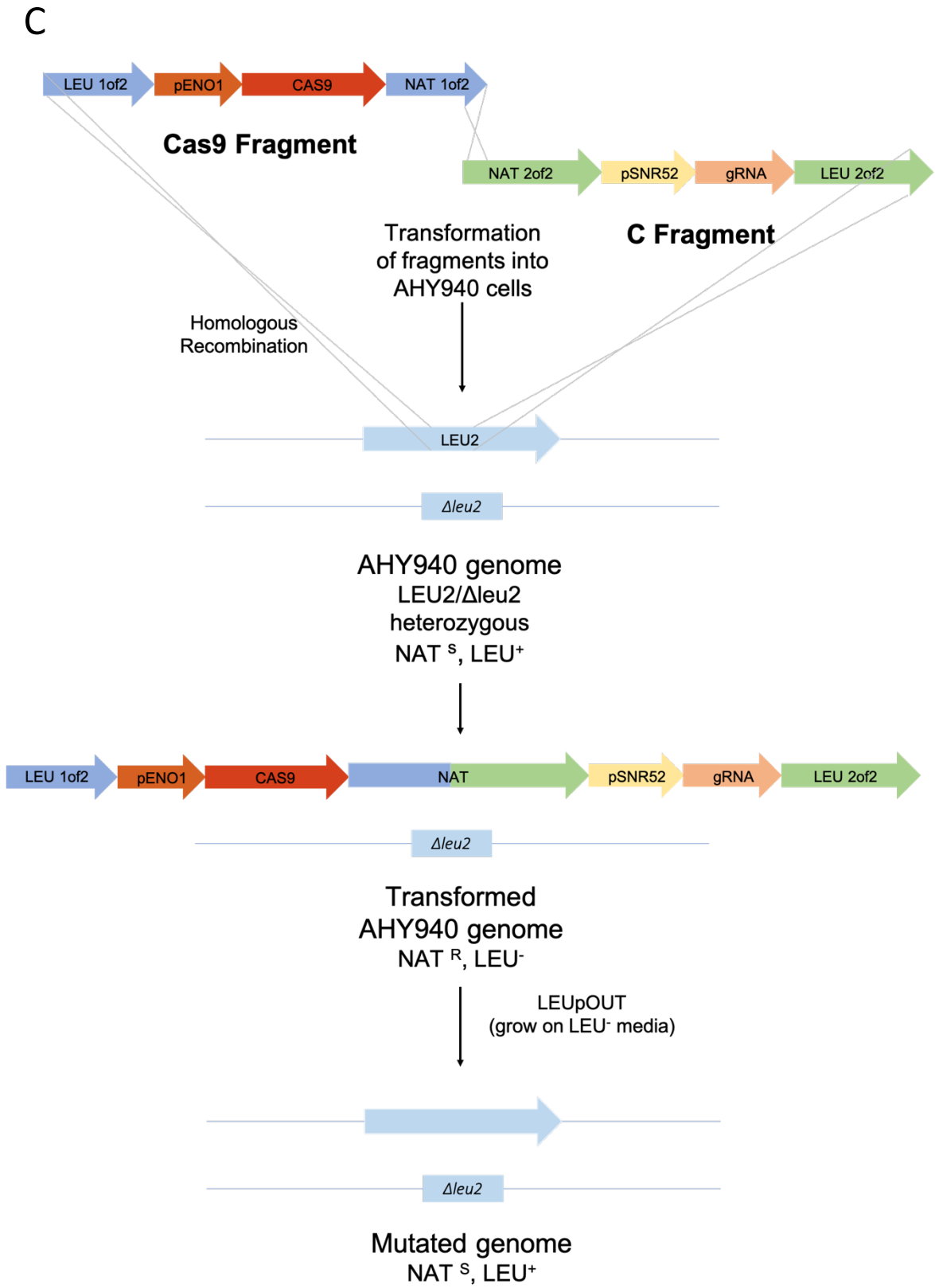


Figure 4.4 Specialised *C. albicans* CRISPR-Cas9 system (C)

Figure 4.4 Specialised *C. albicans* CRISPR-Cas9 system

The process of using the CRISPR-Cas9 system designed by (Nguyen *et al.*, 2017) for gene editing in *C. albicans*. (A) Plasmids containing fragments that make up the system are either PCR amplified or enzymatically digested to generate A, B and *CAS9* fragments. The oligonucleotide primer names are shown, and sequences can be found in Table 2.2. The resulting fragments are shown, with the genetic information they contain. (B) A and B fragments are ligated via PCR to yield the gRNA cassette (C fragment). (C) *CAS9* cassette and gRNA cassette are co-transformed into *C. albicans* strain AHY940 (*LEU2/Δleu2* heterozygous mutant) as two linear DNA segments. The two cassettes homologously recombine via two halves of the NAT cassette, to generate an intact antibiotic resistance marker. The construct then inserts into the one intact *LEU2* locus in the chromosome, via the two sections of *LEU2* homology found at the 5´ and 3´ ends of the *CAS9* and gRNA cassettes respectively.

fragments required to generate the cassettes are amplified using PCR or excised using restriction endonuclease digestion. The gRNA cassette, also called the C fragment, is constructed from the stitching together of an A fragment containing half of the NAT gene and the *SNR52* promoter (an RNA Polymerase III promoter), and a B fragment containing the unique gRNA sequence and one half of the *LEU2* gene. Customisation of the gRNA fragment is achieved through PCR amplification of the B fragment from its host plasmid using an oligonucleotide with a unique target-specific 20 nucleotide segment at the 5´ end. Using this method, the *SNR52* promoter then drives expression of the gRNA *in vivo*. The *CAS9* fragment is supplied on another plasmid (pADH137) and released by enzymatic digestion. The *CAS9* fragment contains the other halves of the *LEU2* gene and NAT cassette described above. The gRNA fragment and the *CAS9* fragment are transformed into the host strain as separate linear fragments, where they homologously recombine due to homology within the NAT cassette and insert the CRISPR system into the *LEU2* locus, disrupting the one intact *LEU2* gene, thus resulting into a *leu*$^-$ phenotype. Once transformants with the correct insertion have been confirmed, the cassette can be removed via the LEUpOUT method (Nguyen *et al.*, 2017), resulting in a nourseothricin sensitive strain, ready to use for further genetic manipulation if required.

4.1.7 Aims of this work

This aim of the work described in this section of the thesis was to successfully create a *TLO* null mutant strain of *C. albicans*, a task which before the advent of CRISPR-Cas9 mutagenesis was deemed too difficult to attempt. The generation of this strain will not only provide a *TLO* null background for further genetic manipulation but will also allow phenotypic analysis of the strain to uncover the consequences of deleting the *TLO* genes.

4.2 Materials and Methods

4.2.1 CRISPR-Cas9 mutagenesis – deletion of TLO *gene family*

4.2.1.1 Design and construction of CRISPR-Cas9 oligonucleotides

The *Candida albicans* CRISPR system described by Nyguen *et al.* in 2017 was used to delete all 14 TLO genes from the parental AHY940 strain, which is a *LEU/Δleu* derivative of the SC5314 lab strain. The conserved Med2 region of the *TLO* genes was selected for the generation of a guide sequence, to enable knocking out both copies of each *TLO* gene in one step. The 20 bp target region, TTTTTGGATTAGTTCGTTGT, was entered into the gRNA oligonucleotide calculator, found in the supplementary material of the publication (Nguyen *et al.*, 2017) to generate the sequence of the full oligonucleotide to be synthesised. This sequence can be found in Table 2.2. A breakdown of the location of the *TLO* genes, where the gRNA binds, and the size of the genes before and after deletion can be found in Table 4.1.

The repair template was designed to have homology to the remaining segment of the Med2 domain and some homology to the variable CTD. The repair template was created by designing two overlapping oligonucleotides, one with the Med2 homology (64 bp in length) and the other with the CTD homology (79 bp in length), which share 20 bp of homology. These oligonucleotides were used in a routine PCR reaction with no template DNA to amplify the primer dimer (123 bp in length). Introduction of this template enables repair of the dsDNA breaks introduced by the Cas9 enzyme at the target sites and creates a truncated ORF for each *TLO* gene.

A positive control was also used during this process to ensure the CRISPR system worked as it should. This control entailed the use of a gRNA targeting the *ADE2* locus based on the 20 bp target sequence, TCTATAGTACAGATGCCAAG, and a repair template that would result in non-functioning copies of this gene. The disruption of *ADE2* leads to red pigmentation of colonies grown on adenine depriving media. Sequences for these oligonucleotides can be found in Table 2.2.

4.2.1.2 Generation of the unique gRNA expression cassette

The “A” fragment, made up of the second half of the nourseothricin resistance gene and the pSNR52 promoter, was amplified from the pADH110 plasmid via PCR using the

TLO Name	Clade	Chromosome	Position	Orientation	gRNA binding site	Size (bp)	Size after deletion
1	α	R	9111-9863	→	9251-9232	753	207
2	β	R	2285377-2286198	←	2286058-2286077	822	215
3	α	1	10718-11485	→	10858-10839	768	206
34	α	1	1292811-1293806	→	1293173-1293154	996	428
4	γ	1	3187464-3187887	←	3187845-3187864	424	unknown
5	γ	2	4248-4778	→	4388-4369	531	384
7	γ	3	13756-14265	→	13896-13877	510	366
8	γ	3	1787576-1788085	←	1787974-1787993	510	363
9	α	4	983-1660	→	1123-1104	678	206
10	α	4	1597159-1597812	←	1597672-1579691	759	199
11	γ	5	1918-2427	→	2058-2039	510	363
12	α	5	1182110-1182868	←	1182728-1182747	759	206
13	γ	6	5545-6069	→	5685-5666	525	377
16	γ	7	942825-943352	←	943212-943231	528	381

Table 4.1 *Candida albicans TLO* genes, CRISPR guide binding locations and sizes before and after CRISPR-Cas9 truncation

The table above lists the name (number) and clade for each *TLO* gene, and the chromosomal coordinates, transcriptional orientation is also listed. The location where the gRNA binds during CRISPR-Cas9 mutagenesis is given, as well as the size of the gene before and after truncation.

AHO1096 and AHO1098 oligonucleotides. Reactions were carried out in 100 µl volumes, with 0.5 U Phusion™ High-Fidelity DNA Polymerase (Thermo Fisher Scientific), 1X Phusion High-Fidelity buffer, dNTPs (200 µM each), 0.5 µM of each oligonucleotide primer (AHO1096 and AHO1098) and 1 ng of template DNA (pADH110). Reaction conditions were as follows: initial denaturing 98 °C for 30 seconds, denaturation at 98 °C for 20 seconds, annealing at 58 °C for 20 seconds, amplification at 72 °C for 30 seconds, cycling between the last three steps for 30 cycles, with a hold at 4 °C until retrieval and storage at 4 °C.

The "B" fragment, comprised of the gRNA sequence and the second half of the *LEU2* gene, was amplified from the pADH119 plasmid via a touchdown PCR using the custom gRNA oligonucleotide described in Section 4.2.1 and the AHO1097 oligonucleotide. Reactions were carried out in 20 µl volumes, with 0.2 U Phusion™ High-Fidelity DNA Polymerase, 1X Phusion HF buffer, dNTPs (200 µM each), 0.5 µM of each oligonucleotide primer (AHO1097 and custom gRNA oligonucleotide) and 0.2 ng of template DNA (pADH119). Reaction conditions were as follows: initial denaturation at 98 °C for 30 seconds, denaturation at 98 °C for 20 seconds, annealing at 65 °C for 20 seconds, amplification at 72 °C for 30 seconds, repeating the last three steps for a total of 10 cycles, reducing the annealing temperature by 1 °C each cycle. This was immediately followed by 25 cycles at the following: denaturing at 98 °C for 20 seconds, annealing at 55 °C for 20 seconds and amplification at 72 °C for 30 seconds. This was then held at 4 °C until retrieval and storage at 4 °C.

The "C" fragment, the gRNA expression cassette, was generated by a fusion of the A and B fragments, with subsequent amplification. The protocols outlined above created enough A and B template for around 75 and 10 stitching reactions, respectively. The C fragment stitching reactions were carried out in 99 µl volumes, with 1 µl of amplification oligonucleotides being added after the initial stitching reaction. The following concentrations of reagents were calculated for 100 µl volumes, as is the total volume for the final step of the procedure: 0.5 U Phusion™ High-Fidelity DNA Polymerase, 1X Phusion HF buffer, dNTPs (200 µM each) with 1 µl each of the A and B fragments. Stitching PCR conditions were as follows: initial denaturation at 98 °C for 30 seconds, denaturation at 98 °C for 20 seconds, annealing at 58 °C for 20 seconds, amplification at 72 °C for 1 min, repeating the last three steps for a total of 5 cycles, then holding at 4 °C until retrieval. After the stitching PCR, amplification oligonucleotides, AHO1237 and AHO1238 were added to a final concentration of 0.5 µM. Amplification reaction

conditions were as follows: initial denaturation at 98 °C for 30 seconds, denaturation at 98 °C for 20 seconds, annealing at 66 °C for 20 seconds, amplification at 72 °C for 1 min, repeating the last three steps for a total of 30 cycles, with holding at 4 °C until retrieval and storage at 4 °C.

The C fragment was then cleaned up as in Section 2.3.7 and the concentration of DNA was quantified as per Section 2.3.3. Generation of the A and B fragments from their respective plasmids can be seen in Figure 4.4 (A), and a representation of the stitching reaction to form the C fragment can be found in Figure 4.4 (B).

4.2.1.5 Generation of the Cas9 expression cassette

The Cas9 cassette containing the first half of the *LEU2* gene, the *CAS9* gene under the *ENO1* promoter, and the first half of the nourseothricin resistance gene, was released from the pADH137 plasmid by digestion with *Mss*I. This reaction was performed in a 10 µl volume, with 2.5 U *Mss*I enzyme, 1X FastDigest buffer and 2 µg of plasmid DNA. This reaction was incubated at 37 °C for a minimum of 30 min. Figure 4.4 (A) shows the release of the Cas9 fragment from the pADH137 plasmid.

The sizes of the products were confirmed by agarose gel electrophoresis (see Section 2.3.4), and the rest of the digestion mix was cleaned and the concentration of DNA determined as above.

4.2.1.6 Transformation of the CRISPR-Cas9 system into *Candida albicans* AHY940

Transformation of the CRISPR-Cas9 system into the *C. albicans* strain AHY940 was performed by electroporation, as described in Section 2.3.8. The gRNA expression cassette (C fragment), the Cas9 expression cassette and the repair template were introduced to the cells in a single transformation reaction. Figure 4.4 (C) depicts the events that occur *in vivo* after transformation of the fragments into the yeast cells. After electroporation the cells were plated on YEPD+NAT200 to select for positive transformants. The positive control, *ADE2* disruption, produced red tinted colonies in the event of a successful transformation. Possible transformants were subcultured onto YEPD+NAT200 plates for further testing.

4.2.1.7 Confirmation of deletions

Mutants were confirmed by routine PCR (see Section 2.3.6) using a *TLO* specific oligonucleotide primer and the pan *TLO* oligonucleotide primer. These primers are

detailed in Table 2.2. The successful deletion of each *TLO* gene was confirmed by a size difference between the WT PCR product and the mutant PCR product, with the mutant product being smaller due to the truncation of the gene by the repair template. Details on the oligonucleotide primer binding locations for each *TLO* gene, as well as PCR product sizes before and after deletion can be found in Table 4.2. Deletions were also confirmed using a whole genome sequencing approach, described in Section 4.2.5.

4.2.1.8 Recycling of the CRISPR system from the *LEU2* locus

After confirmation of the cassette insertions, the CRISPR system required removal from the *LEU2* locus, which restores nourseothricin sensitivity and allows for further genetic manipulation of confirmed deletion mutants. This was achieved by streaking the strains onto YNB without amino acids or ammonium sulphate (leucine negative media) and incubating for 2-3 days at 30 °C. Colonies growing on these plates were then patched on to YEPD agar plates with and without nourseothricin (400 µg/ml). Colonies which grew on the leucine negative plate and exhibited nourseothricin sensitivity were deemed to have lost the CRISPR system. This was confirmed by two rounds of PCR, one to confirm the loss of the system, and another PCR to confirm the mutant genotype (as in Section 4.2.1.7). The AHO1237 and AHO1238 oligonucleotides (used initially to amplify the C fragment), were used to test for the presence of the CRISPR system, with a lack of product indicating successful recycling of the cassette.

4.2.3 CRISPR-Cas9 mutagenesis – depletion of TLOs

The Vyas CRISPR system (Vyas *et al.*, 2015) was used to attempt to knock out the *TLO* gene family in *C. albicans* prior to the Hernday system. This method used SC5314 as the parent strain. Two separate guides were used to attempt to knock out the *TLO* family, neither were successful in generating full knockouts, however they did generate *TLO* depleted strains. Guide sequence 1 (5′-AATGATGCAGAGTGGTGTCT-3′) knocked out all *TLOs* except *TLOγ5*, and guide sequence 2 (5′-ATTGAAAGTCAAAGAAGAAG-3′) knocked out all *TLO*s except *TLOβ2* and *TLOα10*.

4.2.4 Phenotypic analysis

Phenotypic analysis of the confirmed *ΔΔtlo* mutants was carried out as described in Section 2.6. The AHY940 was used as the WT control. Two *ΔΔtlo* mutants arising from

TLO Name	Pan TLO Primer binding site	Chromosome arm specific binding site	PCR Product size	PCR Product size after deletion	Difference
1	9154 - 9169	11116 - 11096	1962	1416	546
2	2286155 - 2286140	2284877 - 2284896	1278	671	607
3	10761 - 10776	12413 - 12393	1651	1089	562
34	*TLO34 F binds: 1292671 - 1292690*	*TLO34 R binds: 1293937 - 1293810*	1166	598	568
4	3187554 - 3187571	3185619 - 3185638	1952	~ 1000	~ 900
5	4291 - 4306	6766 - 6747	2475	2328	147
7	13799 - 13814	16169 - 16150	2370	2230	140
8	1788071 - 1788056	1786004 - 1786023	2067	1920	147
9	1026 - 1041	3198 - 3178	2172	1966	472
10	1597769 - 1597754	1596722 - 1596741	1047	600	447
11	1961 - 1976	4076 - 4057	2115	1968	147
12	1182825 - 1182810	1180630 - 1180648	2195	1642	553
13	5588 - 5603	7415 - 7396	1827	1680	147
16	943309 - 943294	941132 - 941151	2177	2030	147

Table 4.2 *TLO* gene PCR primer binding sites and product sizes before and after CRISPR-Cas9 truncation

two different transformation events were analysed, to confirm phenotypes were as a result of the *TLO* deletions. A *ΔΔmed3* mutant of SC5314 was used as another control strain. This strain still has the full complement of *TLO*s, however, without Med3 it is thought that the Tlo proteins cannot be incorporated into Mediator.

4.2.5 Whole Genome Sequencing analysis

4.2.5.1 Sequencing and initial clean-up

Oxford Nanopore Technology was used to sequence the whole genome of the *ΔΔtlo* mutant of *C. albicans* generated. The sequencing itself and some initial clean-up of the data was performed by a former postdoc in the lab, Dr Peter Flanagan. This resulted in FASTQ files, the analysis of which was performed and is detailed below. A 1D ligation sequencing kit (SQK-LSK109) was used in conjunction with a barcoding kit (EXP-NBD103) to sequence several samples at once. Of these, one was the *ΔΔtlo* mutant with the CRISPR-Cas9 cassette recycled, and another was the second *ΔΔtlo* mutant, which still had the CRISPR-Cas9 cassette incorporated into the genome. Albacore (ONT) was used to basecall the samples, and then Porechop (Wick *et al.*, 2017) was used to demultiplex the reads and trim adaptors from the barcoding kit. The output FASTQ files were then analysed as part of this work by the author.

4.2.5.2 Reference genome

The ONT reads were aligned to *C. albicans* SC5314 reference genome ASM18296v3, which was acquired from the NCBI database (most recently on 26th October 2020) (Jones *et al.,* 2004). The reference genome was downloaded from the NCBI website in a FASTA (FNA) format. A General Feature Format file (GFF) was also downloaded from the NCBI website. This file describes the genomic features of the DNA.

4.2.5.3 NanoPack – quality checking and filtering data

The NanoPack tool suite (De Coster *et al.*, 2018) was used for quality checking the data, and for filtering the data based on the results of quality checking.

NanoQC, a part of NanoPack, was used to generate plots to investigate the quality of the base calls at the start and end of reads. These plots were then used to determine the number of bases that should be trimmed from the start of each read before aligning. The

following command was used to generate the .html file from NanoQC which was then visualised through Bokeh (bokeh.org) from which images were downloaded.

```
nanoqc –o QC reads.fastq
```

NanoPlot was employed to generate a statistical report about the reads in the fastq file. This tool also generates several plots that give a summary of the read lengths in the sample. NanoPlot was used to determine a suitable quality score to use as a cut-off point for filtering the reads before alignment.

```
nanoplot –o Plots ––N50 ––fastq reads.fastq
```

NanoFilt, another NanoPack tool, was used to filter the data based on q score, eliminating any read with a q score of less than 10, and eliminating any read with a length of less than 500 It also cropped the first 50 and last 50 base calls from each read.

```
nanofilt –q 10 –l 500 ––headcrop 50 ––tailcrop 50 reads.fastq > filteredreads.fastq
```

4.2.5.4 Estimating coverage

The theoretical coverage of the run was calculated for both the filtered and unfiltered FASTQ files. This was estimated based on the following equation:

Number of reads (mean read length / genome size)

Genome size of *Candida albicans* used was 14.7 Mb. The number of reads and the mean read length were obtained by running the NanoPlot tool (above).

4.2.5.5 Burrows-Wheeler Aligner – aligning read to the reference genome

The Burrows-Wheeler Aligner (BWA) (Li, 2013) was used to align the FASTQ files to the reference genomes. Before the alignment, both the reference FNA file and the FASTQ reads files were indexed with bwa index:

```
bwa index reference.fna OR bwa index filteredreads.fastq
```

After indexing, the reads were aligned to the reference genome using bwa mem, and the result outputted into a Sequence/Alignment Map file (SAM).

```
bwa mem reference.fna filteredreads.fastq > alignment.sam
```

4.2.5.6 SAMtools – processing the alignment file

Samtools (Li *et al.*, 2009) was used to convert the SAM file into a compressed SAM file (BAM), and then to sort the BAM file and index it.

```
samtools view –O bam –o alignment.bam alignment.sam
```

```
samtools sort –T temp –O bam –o alignment.sorted.bam alignment.bam
```

```
samtools index alignment.sorted.bam
```

4.2.5.7 Integrative Geonomics Viewer – visualising the alignment

Integrative Geonomics Viewer (IGV) software (Robinson *et al.*, 2011) was used to visualise the generated alignment. The reference genome and reads were uploaded and the regions of the *TLO* genes were viewed to confirm deletion/disruption of the alignment at these regions.

4.2.5.8 Determining and plotting depth of coverage per chromosome

To determine if there were any major and obvious changes in karyotype of the *ΔΔtlo* strain-, the coverage over each chromosome was determined. If the coverage over any chromosome was significantly higher than others it could indicate aneuploidy in the sample. This can also be used to check that coverage over a chromosome was even.

The depth of coverage over every position was determined with SAMtools using the following command, which outputted this information into its own coverage file:

```
samtools depth alignment.sorted.bam > alignment.coverage
```

This coverage file was then split into coverage files for each individual chromosome:

```
awk '$1 == "chromosome_name" {print $0} alignment.coverage > chromosome_name.coverage
```

These coverage files were then plotted using R (R Core Team).

4.2.6 Examining chromosome structure via CHEF gel

To examine chromosome structure of the mutants, the strains were sent to the lab of Dr Alessia Buscaino at the University of Kent, where clamped homogenous electric fields (CHEF) gel electrophoresis was carried out.

4.3 Results

4.3.1 Confirmation of deletion of TLO *genes in* C. albicans *strain AHY940*

Following transformation with the *TLO* gRNA expression cassette routine PCR was performed to confirm the deletion of all of the *TLO* genes in the knockout strains. The primers used to amplify each *TLO* gene were a pan-*TLO* oligonucleotide primer with homology to the conserved *MED2* domain common among all *TLO* genes and a chromosome arm-specific primer, Figure 4.5 (A). Due to the mid-chromosomal location of the *TLO34* locus, a distinct oligonucleotide primer set was used to target this locus. A substantial size difference in the products of these PCR reactions indicated a truncation of both copies of the *TLO* gene, meaning that the deletion via CRISPR Cas9 mutagenesis, and subsequent repair via the repair template was successful.

Details of the CRISPR gRNA binding site relative to each *TLO* locus, along with predicted PCR sizes before and after mutagenesis can be found in Table 4.2. An image of the PCR products run on an agarose gel can be seen in Figure 4.5 (B&C).

Many colonies screened had some of the *TLO* genes still intact with only some being successfully knocked out. Efficiency in obtaining complete *TLO* knock outs was not high. In one transformation around 35 colonies were screened before a null mutant was obtained, and in another approximately 25 colonies were screened, an efficiency rate of 2.85% and 4% respectively. Mean efficiency was 3.3%. Two replicates of the CRISPR Cas9 mutagenesis reaction were carried out, and from each reaction a mutant strain was confirmed to have all the *TLO* genes deleted. Both of these strains were used for phenotypic analysis, to confirm that any phenotypic changes seen were due to the deletion of the *TLO* gene family. The strain referred to in this work as the *ΔΔtlo* strain was generated first, and the *ΔΔtlo*(#2) strain was generated from a second experiment.

4.3.2 Confirmation of recycling of the CRISPR Cas9 cassette from mutant strains

In order to introduce any further mutations into the *ΔΔtlo* strain, the CRISPR Cas9 cassette, including the nourseothricin resistance marker must be removed from the chromosome. The CRISPR Cas9 mutagenesis system used here (Nguyen *et al.*, 2017),

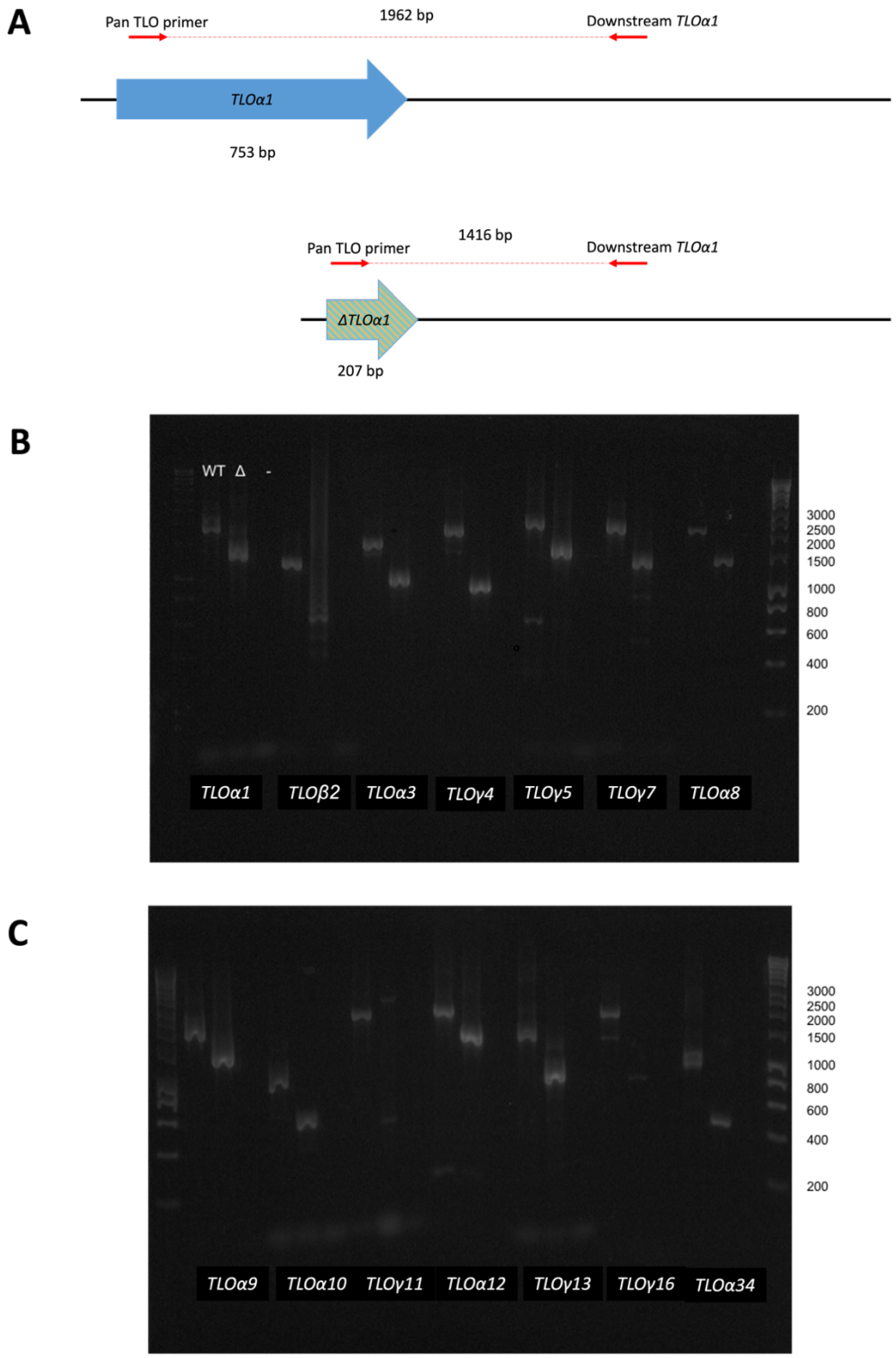


Figure 4.5 PCR confirmation of *TLO* deletions

(A) The loci where the *TLO* genes are located on the chromosomes were amplified with a pan-*TLO* oligonucleotide primer and a chromosome arm-specific oligonucleotide primer for each *TLO.* Successful deletion of the *TLO*s were confirmed by a smaller size in the mutant strain than the WT strain for each *TLO* product. This is due to the deletion of the *TLO*s via CRISPR and the subsequent repair creating a truncated ORF. TLOα1 truncation is used as an example. (B&C) The sizes of the PCR products confirmed that each of the TLO genes was truncated as expected in the mutant. Due to its position in the middle of chromosome 1, two specific oligonucleotide primers were used to amplify the *TLO34* locus.

allows for easy, rapid recycling of the cassette through use of the LEUpOUT system, which is detailed in Section 4.2.1.8. Loss of the CRISPR Cas9 cassette was confirmed in one colony, and this was used as the primary *ΔΔtlo* mutant going forward. This is the strain referred to as *ΔΔtlo*, the *ΔΔtlo* (#2) strain still contains the resistance cassette. All phenotypes described below were shared by the mutants, with and without the CRISPR Cas9 cassette.

4.3.3 Phenotypic analysis of C. albicans ΔΔtlo *strains*

In this section, the strain referred to as *ΔΔtlo* is the mutant with all the *TLO* genes knocked out, and which has had the CRISPR Cas9 cassette recycled. The strain referred to as *ΔΔtlo* (#2) is the strain from the second transformation event, with all the *TLO* genes knocked out. However, this strain retains its CRISPR Cas9 cassette and nourseothricin resistance marker. The strain referred to as *ΔΔmed3* is a mutant strain derived from an SC5314 parent, which is homozygous null for the *MED3* gene, but has all the *TLO* genes intact. A description of all strains and mutants can be found in Table 2.1.

Results of the phenotypic analysis of the *TLO* depleted strains generated through the Vyas CRISPR method can be found in Appendix 1.

4.3.3.1 Colony and cellular morphology

The morphology of the strains was examined both at a colony level and at a cellular level to identify any significant differences between the mutant and WT strains. After 48 h growth on solid YEPD at 37 °C, colonies of the *ΔΔtlo* mutants appeared macroscopically similar in morphology to the AHY940 WT, but slightly smaller in size. Images of the colony morphologies can be seen in Figure 4.6. The *ΔΔmed3* strain displayed a more wrinkled morphology than either the WT or the *ΔΔtlo* strains.

When cells from overnight cultures grown in YEPD broth were observed under the microscope, cellular morphology of the *ΔΔtlo* strains appeared dramatically different to the WT strain, with the *ΔΔtlo* strains being much more pseudohyphal than the WT. This can be seen in Figure 4.6. The *ΔΔmed3* strain also displayed this pseudohyphal morphology.

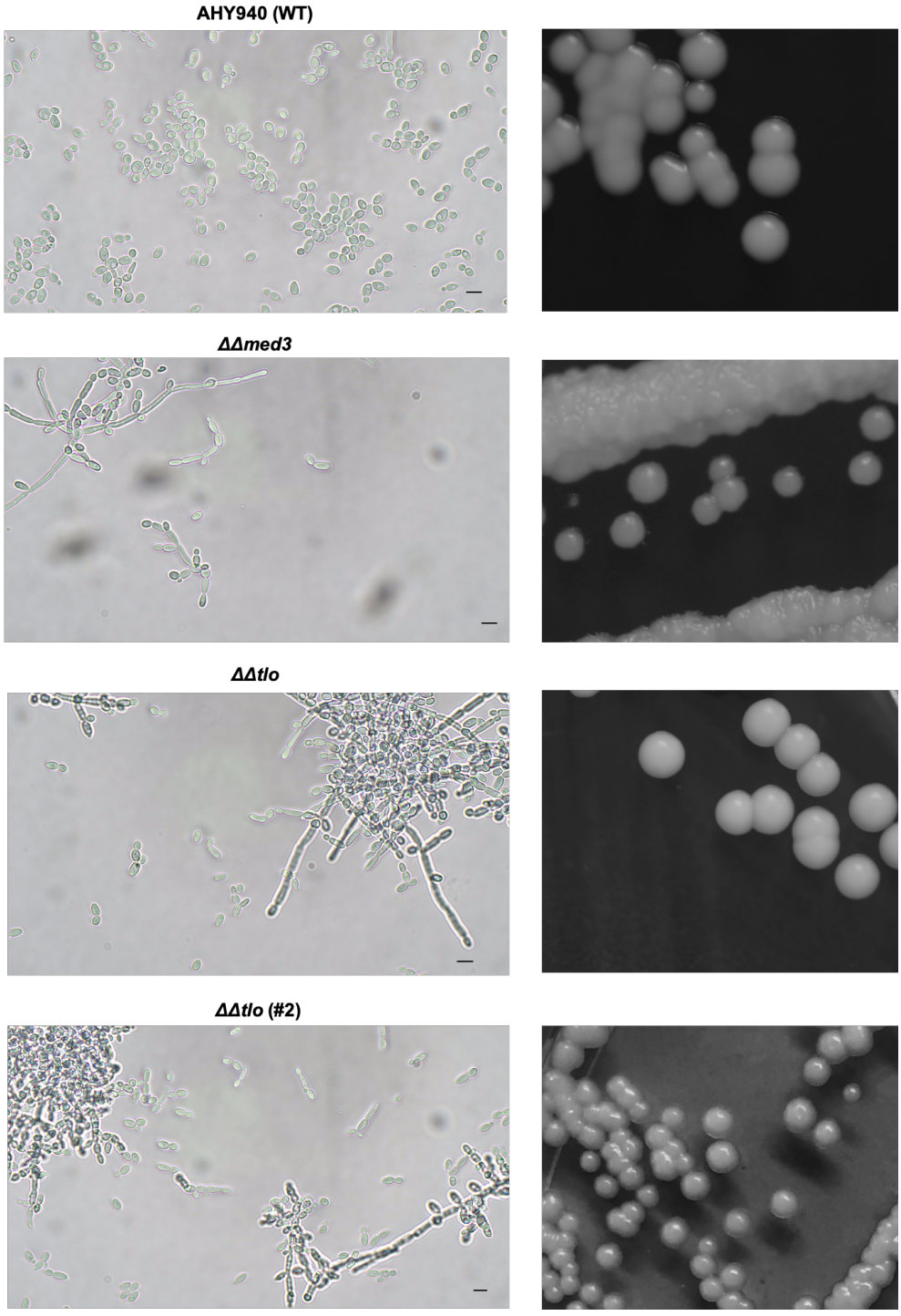


Figure 4.6 Colony and cellular morphology of *C. albicans ΔΔtlo* mutant strains

Colony morphology (left) was photographed after 48h of growth on YEPD agar at 37 °C. Cellular morphology was imaged under X400 magnification after 24h of growth in YEPD liquid media. Scale-bar represents 7 µm.

4.3.3.2 Hyphal induction

The ability of the strains to form true-hyphae in inducing conditions was tested. Hyphal induction with YEPD supplemented with 10% FCS failed to produce true hyphae for any strain apart from the WT AHY940. The other strains maintained their pseudohyphal morphology. Images from these experiments can be seen in Figure 4.7. While the pseudohyphal protrusions did appear to elongate, these were not classed as true hyphae due to their uneven sides and the presence of constrictions along the elongated cells.

Strains were also grown on solid Spider medium for 5 days at either 30 °C or 37 °C to induce hyphal formation, Figure 4.8. When grown at 30 °C the wild type developed hyphae that invaded the agar, however the *ΔΔtlo* and *ΔΔmed3* strains did not show this same level of hyphal growth. The *ΔΔmed3* strain appeared more wrinkled than the *ΔΔtlo* mutants. After growth at 37 °C a much more significant difference was seen between the strains. The wild type showed a large amount of hyphal growth. At this temperature, there appeared to be a difference between hyphal growth of the *ΔΔtlo* and *ΔΔmed3* strains; the *ΔΔmed3* mutant seemed to grow more like the wild type, but the *ΔΔtlo* mutants displayed a greatly reduced amount of hyphal growth and these were the least wrinkled of the strains tested.

4.3.3.3 Chlamydospore formation

The ability of strains to form chlamydospores in inducing conditions was tested by growing the strains on corn meal agar supplemented with 1% Tween 80. Only the WT AHY940 strain was able to form chlamydospores, the *ΔΔtlo* stains and the *ΔΔmed3* strain could not form chlamydospores under these conditions, Figure 4.9.

4.3.3.4 Growth rate analysis

The growth rates of the strains were examined in both YEPD and YEP-Galactose, the results of which can be found in Figure 4.10. In liquid YEPD, the *ΔΔtlo* and *ΔΔmed3* strains all grew significantly slower than the WT. However, it was seen that in YEP-Galactose, while all strains did grow significantly slower than the WT, there was a difference in the growth rates of the *ΔΔtlo* strains and that of the *ΔΔmed3* strain. With both the *ΔΔtlo* strains doubling more rapidly than the *ΔΔmed3* strain.

The ability of strains to grow at various temperatures was examined via spot plate analysis, Figure 4.11. While growth did not seem to be majorly affected by incubation temperature, it did appear that temperature has an effect on morphology. While WT

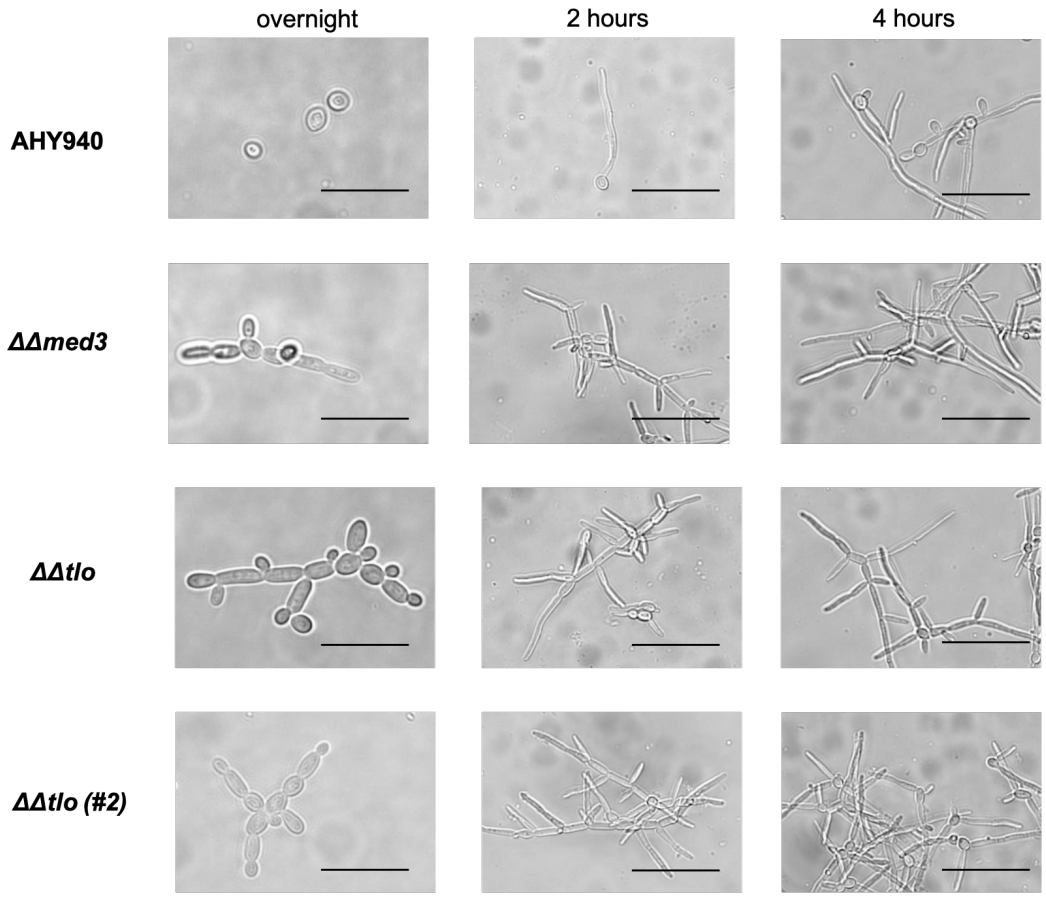


Figure 4.7 *ΔΔtlo* mutant hyphal formation assay in YEPD with 10% Serum

The ability of the strains to form true hyphae under inducing conditions was measured by growing overnight cultures of the strains at 30 °C then standardising to an OD_{600} = 0.2 in YEPD supplemented with 10% FBS in 2 ml volumes in 6 well tissue culture plates. These plates were incubated in a static incubator at 37 °C. Representative images of the cells were taken at various time points. Scale bar equals 7 μm in the overnight panels and 16 μm in the 2- and 4-hour panels.

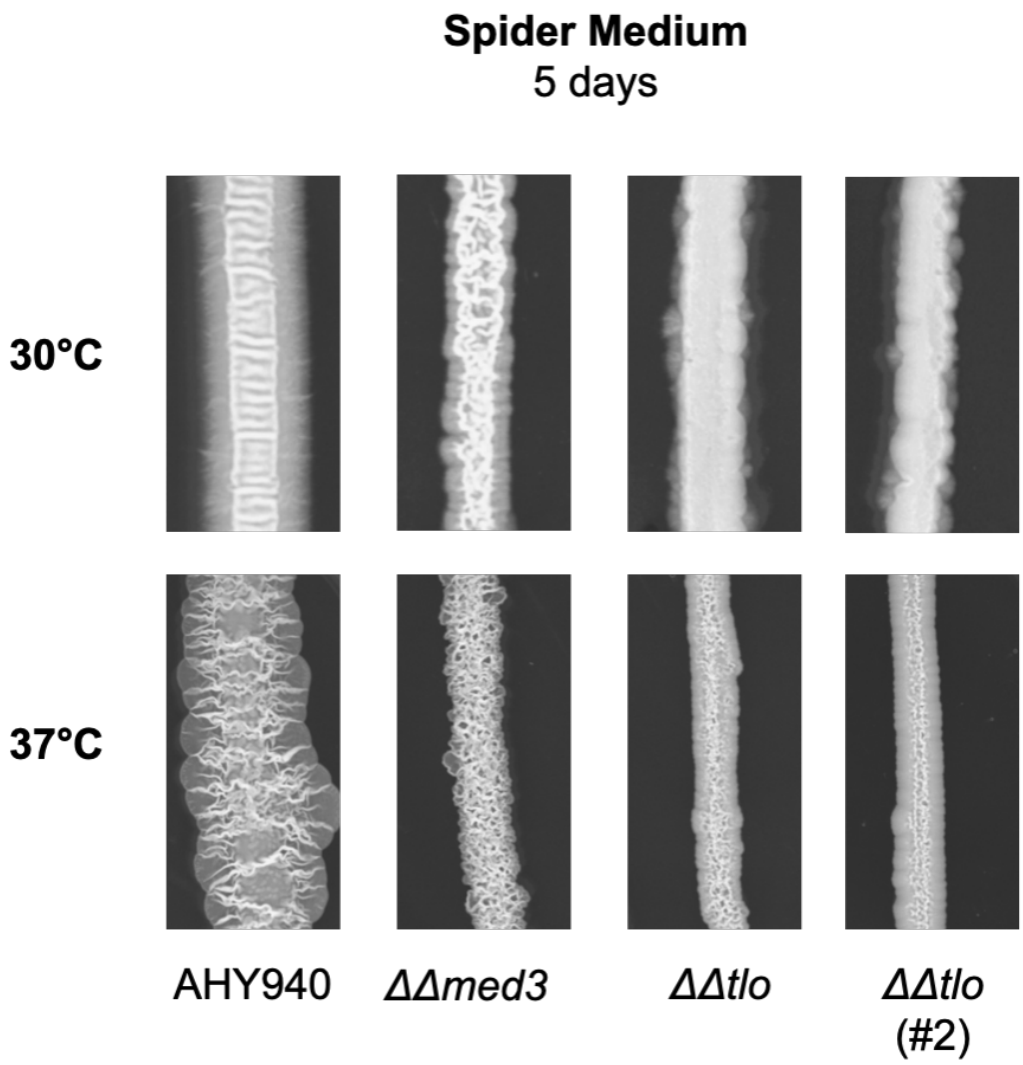


Figure 4.8 Growth of *ΔΔtlo* mutant strains on solid Spider media

The morphology of strains cultured on solid Spider medium was observed by streaking a single colony of each strain onto an agar plate and incubating at either 30 °C or 37 °C for 5 days.

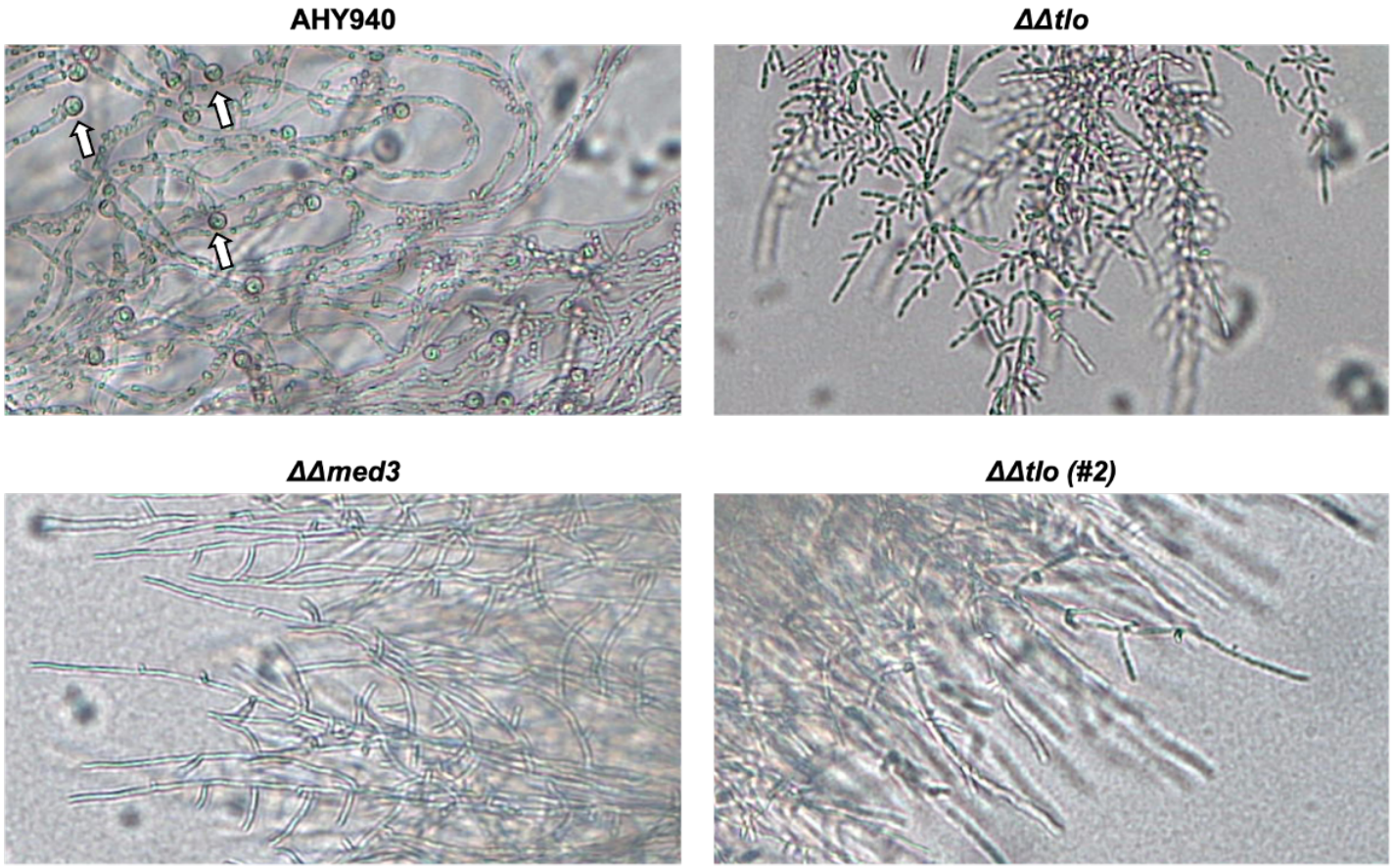
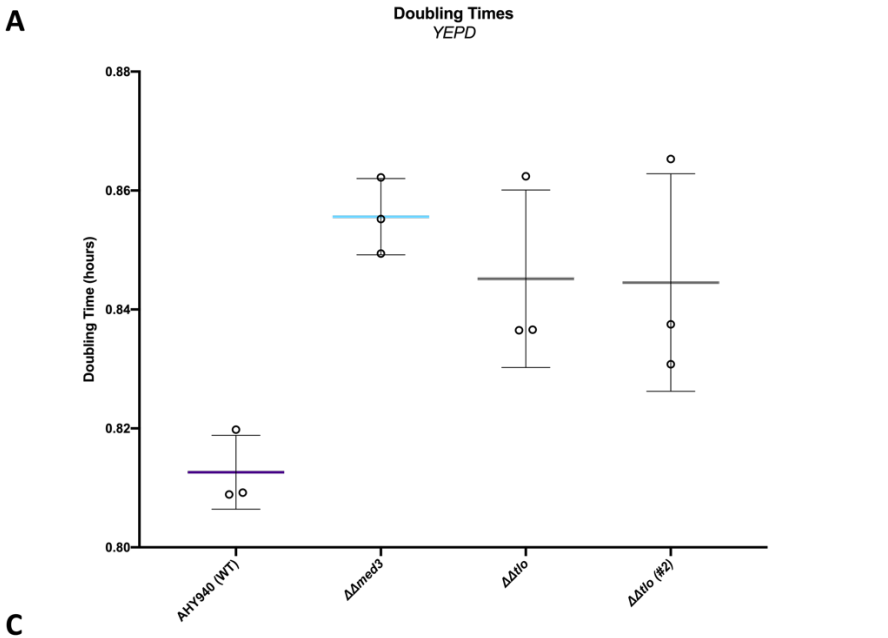
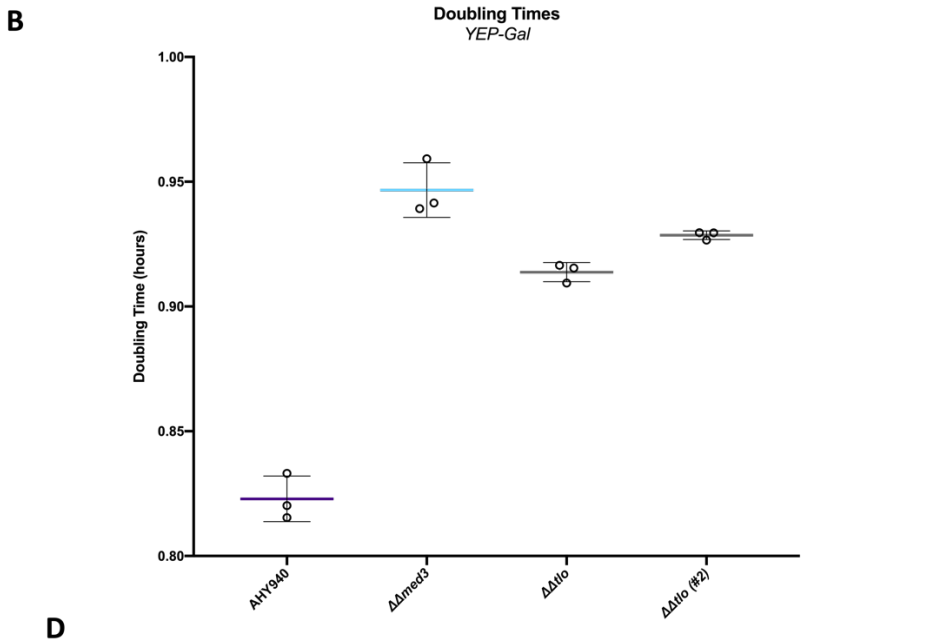


Figure 4.9 Chlamydospore formation in *ΔΔtlo* mutant *C. albicans* strains

Each strain was streaked gently onto corn meal agar supplemented with 1% Tween80, covered with a glass coverslip and incubated in the dark at room temperature for 5 days. Plates were examined under the microscope and images were taken. Representative chlamydospores are indicated by white arrows in the AHY940 panel.





C

YEPD	AHY940	*ΔΔmed3*	*ΔΔtlo*	*ΔΔtlo* (#2)
Mean (h)	0.8126	0.8556	0.8452	0.8445
S.D.	0.00621	0.00641	0.01492	0.01829
Significantly different to AHY940	-	**	*	*
P-value	-	0.008	0.0332	0.0364
Significantly different to ΔΔmed3	**	-	ns	ns
P-value	0.008	-	0.636	0.5976

D

YEP-Gal	AHY940	*ΔΔmed3*	*ΔΔtlo*	*ΔΔtlo* (#2)
Mean (h)	0.8229	0.9466	0.9138	0.9286
S.D.	0.00915	0.01094	0.00382	0.0017
Significantly different to AHY940	-	****	****	****
P-value	-	<0.0001	<0.0001	<0.0001
Significantly different to ΔΔmed3	****	-	**	*
P-value	<0.0001	-	0.0016	0.0433

Figure 4.10 Growth rate analysis of *ΔΔtlo* mutant strains

Doubling time of each strain is represented by horizontal line (hours), with error bars representing standard deviation and symbols representing each of three replicates. (A) Doubling times of strains growing in liquid YEPD at 37 °C at 200 rpm. (B) Doubling times of strains growing in liquid YEP-Galactose at 37 °C at 200 rpm. (C) Statistical analysis of data from strains growing in YEPD. A one-way ANOVA was performed to determine if results were significantly different and a Dunnett's multiple comparisons test was performed to determine which means were significantly different from WT AHY940, and another to determine significant difference from the ΔΔmed3 strain, asterisks represent degrees of significance. (D) As with C, represents statistical analysis of strains grown in YEP-Galactose.

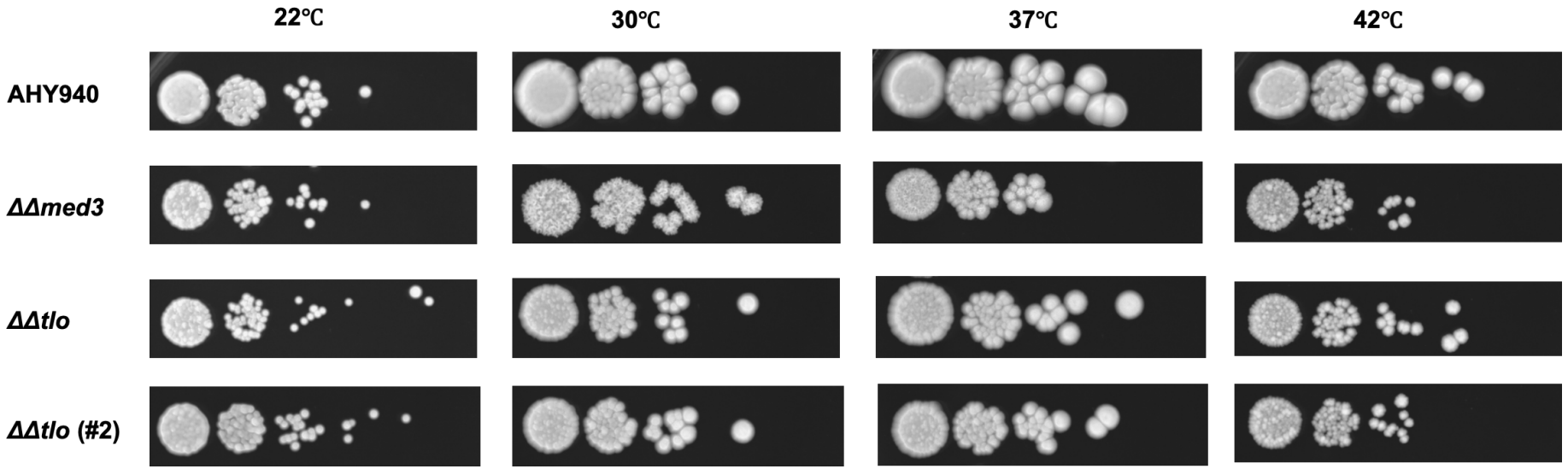


Figure 4.11 Growth of *ΔΔtlo* mutant strains at various temperatures

10-fold serial dilutions (10^6 – 10^1 cfu/ml) were plated in 5 µl volumes onto YEPD agar and incubated for 72 h at 22, 30, 37 or 42 °C before photographing.

morphology appeared similar (smooth) at all temperatures tested, the *ΔΔtlo* strains developed a more irregular and winkled morphology when grown at 42 °C. This also appeared to be the case for the *ΔΔmed3* strain also. At 30 °C, this wrinkled morphology was seen more intensely in the *ΔΔmed3*, however this was not seen in either the WT or *ΔΔtlo* strains.

The ability of strains to grow in the presence of a cellular respiration inhibitor was tested by performing a spot plate assay on YEP-Galactose with and without Antimycin A. Images of strains spotted onto a control YEP-Galactose only plate versus a YEP-Galactose with 1 μg/ml Antimycin A can be seen in Figure 4.12. The *ΔΔtlo* strains and the *ΔΔmed3* strain all grew less well than the AHY940 WT strain on the media containing Antimycin A, indicating a reduced ability to grow when respiration is inhibited. The *ΔΔtlo* and *ΔΔmed3* mutant strains appeared to grow similarly in these conditions, and no differences were observed on Antimycin A. It should be noted however, that the *ΔΔmed3* strain displayed a wrinkled colony morphology when growing on the YEP-Galactose control plates, which is different to the smooth morphology of the *ΔΔtlo* mutant strains. The AHY940 WT strain also displayed a wrinkled morphology on the YEP-Galactose, however, this was much more pronounced than that of the *ΔΔmed3* mutant.

4.3.3.5 Biofilm formation

The ability of strains to form biofilms on plastic surfaces was examined. Spider liquid broth was used as a growth medium in these experiments. Graphical representation of the results at 24 and 48 h can be seen in Figure 4.13. At both 24 and 48 h the *ΔΔtlo* and *ΔΔmed3* mutant strains formed less biofilm than the WT AHY940 strain. There were no differences in biofilm formation between the *ΔΔtlo* strains, or between the *ΔΔtlo* strains and the *ΔΔmed3* mutant.

4.3.3.6 Cell wall perturbing compound spot plate assays

Comparative growth of strains in the presence of cell wall perturbing compounds, Calcofluor White and Congo Red, was determined via spot plate assay, Figure 4.14. When grown on media containing 50 μg/ml Calcofluor white, the *ΔΔtlo* and *ΔΔmed3* appeared more sensitive than the WT AHY940. When the concentration of Calcofluor White was increased to 150 μg/ml, a difference between the *ΔΔtlo* strains and the *ΔΔmed3* mutant became more apparent, with the *ΔΔmed3* strain showing more sensitivity to the compound than the *ΔΔtlo* strains. When cultured on media containing

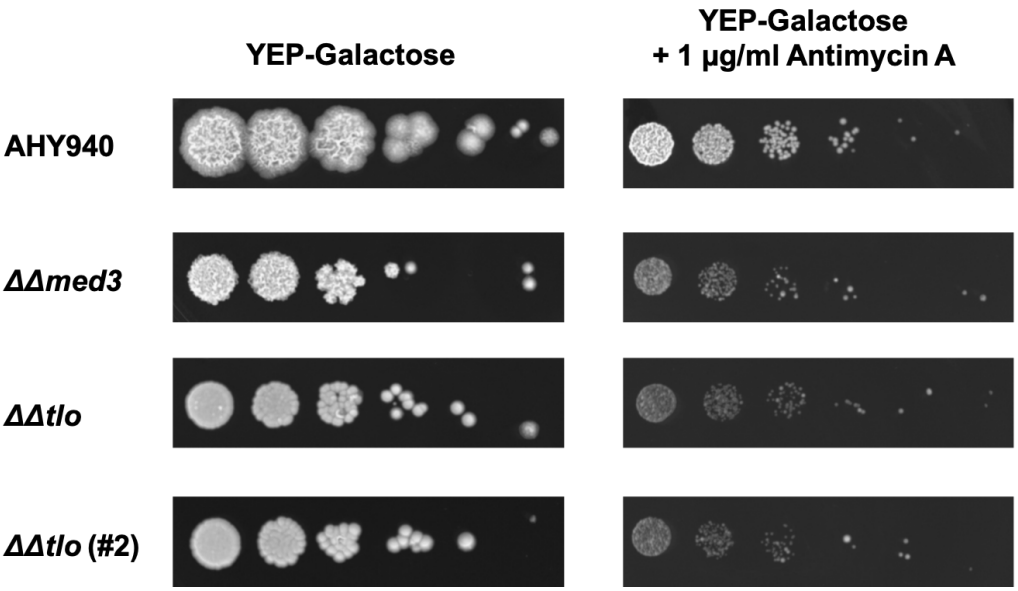
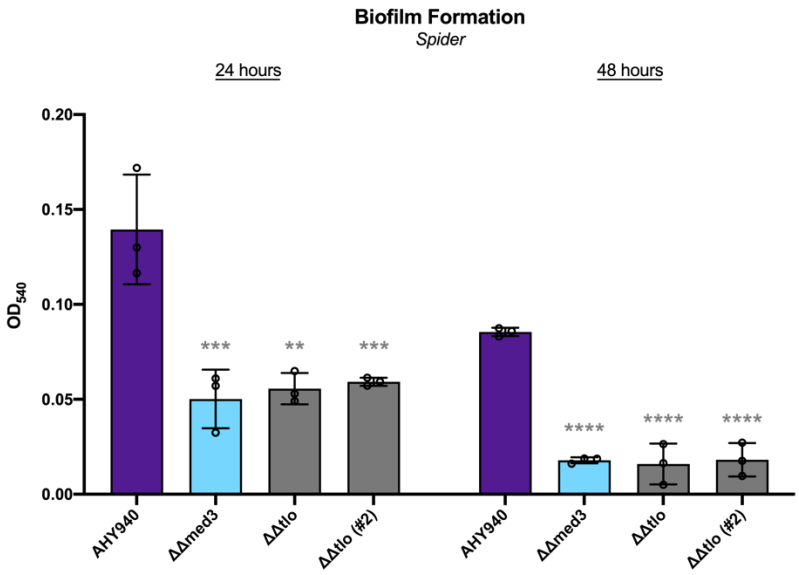


Figure 4.12 Growth of *ΔΔtlo* mutant strains in the presence of Antimycin A

10-fold serial dilutions (10^6 – 10^1 cfu/ml) were plated in 5 µl volumes onto YEP-Galactose agar with and without 1 µg/ml Antimycin A and incubated for 72 h at 37 °C before photographing.

A



B

24 h	AHY940	ΔΔmed3	ΔΔtlo	ΔΔtlo (#2)
Mean (OD_{540})	0.1394	0.05016	0.05560	0.05921
S.D.	0.02888	0.01546	0.008263	0.002100
Significantly different to AHY940	-	***	***	**
P-value	-	0.0005	0.0008	0.0011
Significantly different to ΔΔmed3	***	-	ns	ns
P-value	0.0005	-	0.9590	0.8506

C

48 h	AHY940	ΔΔmed3	ΔΔtlo	ΔΔtlo (#2)
Mean (OD_{540})	0.08549	0.01784	0.01590	0.01811
S.D.	0.002146	0.001540	0.01072	0.008796
Significantly different to AHY940	-	****	****	****
P-value	-	<0.0001	<0.0001	<0.0001
Significantly different to ΔΔmed3	****	-	ns	ns
P-value	<0.0001	-	0.9732	>0.9999

Figure 4.13 *ΔΔtlo* mutant strains' ability to form biofilm on plastic surfaces

The ability of strains to form biofilm on plastic surfaces (plastic 96 well plate) in liquid Spider medium was analysed. Biofilm formation was quantified by staining with crystal violet and measuring absorbance at 540 nm. This was done at 24- and 48-hour time points. (A) Graphical representation of biofilm formation data. Error bars represent standard deviation from the mean, and symbols represent values from three replicates. Asterisks denote statistically significant difference from WT or from *ΔΔmed3* strains. (B) Summary of data and statistical analysis (one-way ANOVA) of 24 h time point data. (C) Summary of data and statistical analysis (one-way ANOVA) of 48 h time point data.

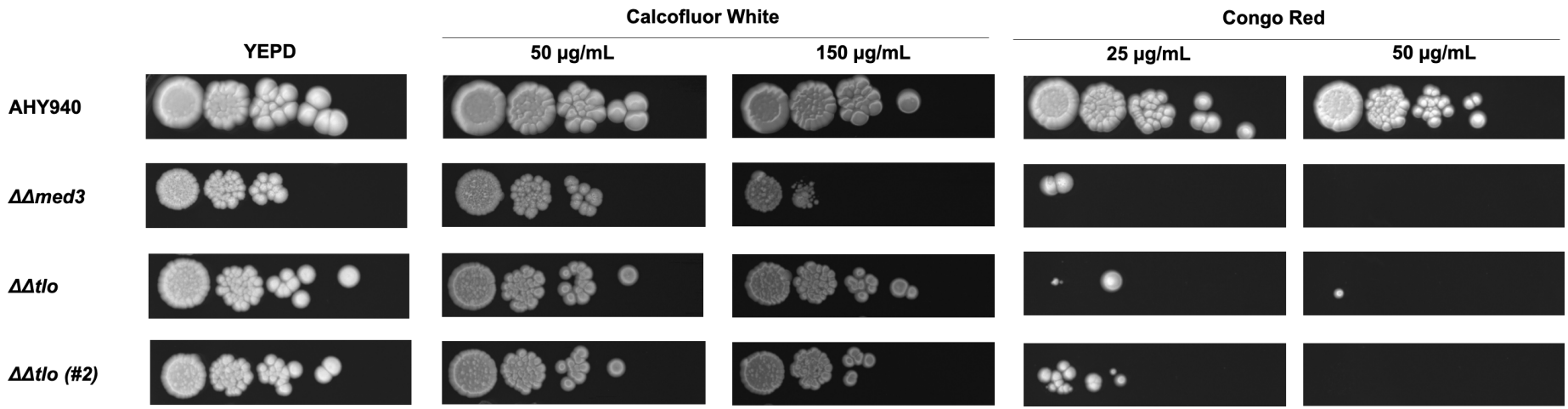


Figure 4.14 Growth of *ΔΔtlo* mutant *C. albicans* strains on media containing cell wall perturbing compounds

10-fold serial dilutions (10^6 – 10^1 cfu/ml) were plated in 5 µl volumes onto YEPD agar containing different concentrations of either Congo Red or Calcofluor White to determine susceptibility of strains to these compounds. Plates were incubated at 37 °C for 72 h before photographing.

Congo Red (25 µg/ml and at 50 µg/ml), growth was dramatically inhibited in the *ΔΔtlo* and *ΔΔmed3* strains compared to that of the WT.

4.3.3.7 Oxidative stress minimum inhibitory concentration assays

Sensitivity of strains to oxidative stress was analysed through microbroth dilution assays with various concentrations of the oxidative stress inducing compounds H_2O_2 and tBOOH. Graphical representation of the IC_{50} for the strains tested can be seen in Figure 4.15. For both compounds, the IC_{50} concentrations for the *ΔΔtlo* and *ΔΔmed3* strains were much lower than that of the WT. There was no significant difference in the IC_{50} concentrations of the *ΔΔtlo* and the *ΔΔmed3* mutants.

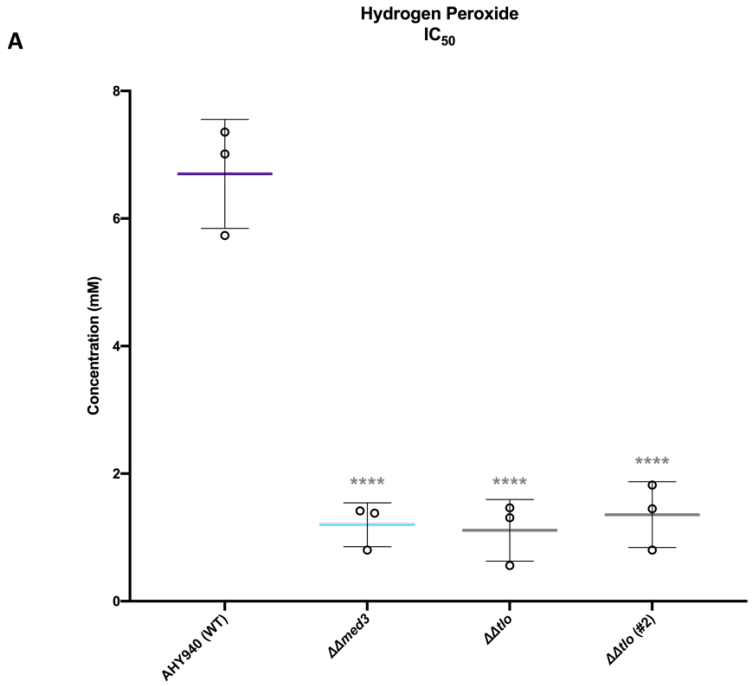
4.3.3.8 *Galleria mellonella* infection model

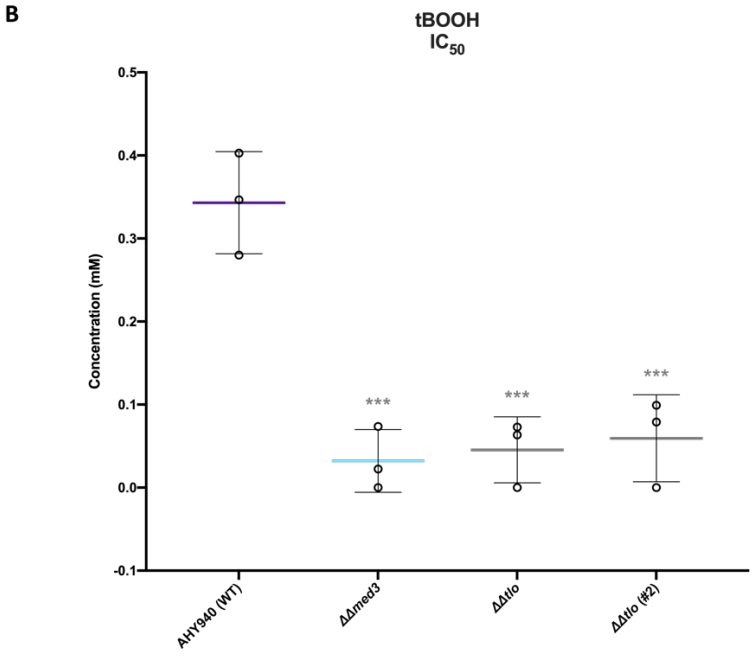
The effect of deletion of the *C. albicans TLO* gene family on virulence in a *Galleria mellonella* infection model was determined, and the results are presented in a Kaplan-Meyer curve which can be found in Figure 4.16. The *ΔΔtlo* mutant strains were significantly less virulent than the WT AHY940, with no significant difference between the two *ΔΔtlo* mutants. The *ΔΔmed3* strain displayed a similar level of virulence as the WT AHY940 strain, and statistically, the *ΔΔmed3* was more virulent than the *ΔΔtlo* strains.

4.3.4 Whole Genome Sequencing to confirm deletion of TLO *gene family*

Whole Genome Sequencing was performed on the *ΔΔtlo* mutants to determine if the *TLO* genes were successfully disrupted. The sequences were filtered for quality and length, then aligned to the reference genome and visualised using IGV. The output of the quality control statistics from the NanoStat tool before and after filtering can be found in Appendix 2.

The location for each *TLO* gene was viewed and the deletions confirmed. These snapshots can be seen in Appx. 3 Figure 1 for the *ΔΔtlo* strain with the CRISPR cassette recycled, while results for the *ΔΔtlo* (#2) strain with the cassette still in place can be found in Appx. Figure 3. The coverage across the genome was plotted for each chromosome and can be seen in Appx. 3 Figure 2 for the *ΔΔtlo* strain, while results for the *ΔΔtlo* (#2) strain can be found in Appx. 3 Figure 4. It can be seen that there is relatively even coverage across the entire length





C

H_2O_2	AHY940	*ΔΔmed3*	*ΔΔtlo*	*ΔΔtlo* (#2)
Mean (mM)	6.699	1.198	1.11	1.356
S.D.	0.8544	0.3445	0.4834	0.5157
Significantly different to AHY940	-	****	****	****
P-value	-	<0.0001	<0.0001	<0.0001
Significantly different to ΔΔmed3	****	-	ns	ns
P-value	<0.0001	-	0.9952	0.974

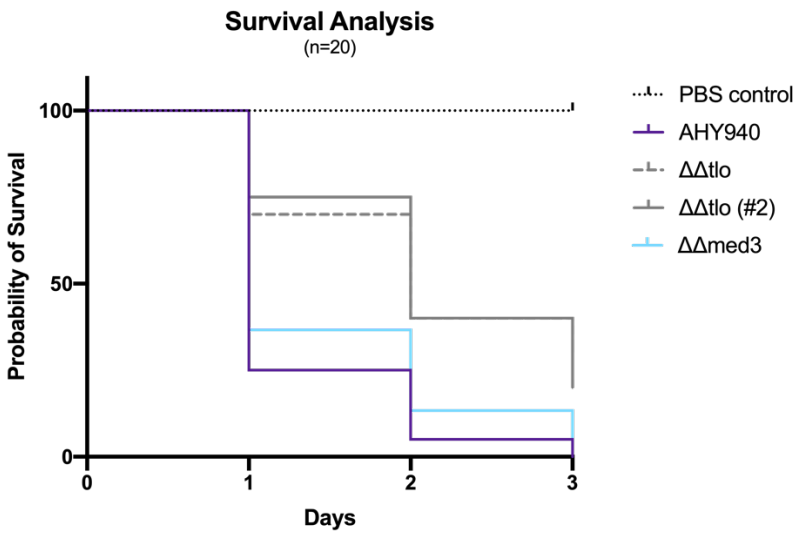
D

tBOOH	AHY940	*ΔΔmed3*	*ΔΔtlo*	*ΔΔtlo* (#2)
Mean (mM)	0.3431	0.03203	0.04549	0.05936
S.D.	0.06152	0.03776	0.03967	0.05239
Significantly different to AHY940	-	***	***	***
P-value	-	0.0002	0.0002	0.0003
Significantly different to ΔΔmed3	***	-	ns	ns
P-value	0.0002	-	0.9731	0.8341

Figure 4.15 Susceptibility of *ΔΔtlo* mutant *C. albicans* strains to oxidative stress

Minimum inhibitory concentration assays were performed by microtiter dilution to determine the IC_{50} (concentration at which growth is inhibited by 50%) of strains to oxidative stress inducing reagents. (A) IC_{50} concentrations of each strain cultured in hydrogen peroxide (H_2O_2), mean is indicated by the horizontal bar, with error bars representing standard deviation and symbols representing each of three replicates. (B) IC_{50} concentrations of strains in tBOOH. (C) Summary of data and statistical analysis (one-way ANOVA) of growth of strains in H_2O_2. (D) Summary of data and statistical analysis of growth of strains in tBOOH.

A



B

	AHY940	*ΔΔmed3*	*ΔΔtlo*	*ΔΔtlo (#2)*
Significantly different to AHY940	-	ns	***	***
P-value	-	0.2909	0.0007	0.0006
Significantly different to ΔΔmed3	ns	-	**	**
P-value	0.2909	-	0.003	0.0024

Figure 4.16 Survival of *Galleria mellonella* infection model after infection with *ΔΔtlo* mutant *C. albicans* strains

Each *Galleria mellonella* larva was infected with 10^6 cells. Inoculation with same volume of PBS was used as a negative control. In total 20 worms were infected with each strain/PBS. (A) Kaplan-Meyer curve of survival of worms over 72 h after infection with *C. albicans* strains (B) Statistical analysis of results. Strains were compared to AHY940 and to the *ΔΔmed3* strain using a Log-rank (Mantel-Cox) test. Degrees of significance are indicated by asterisks, and P-values are given.

of each chromosome, with the exception of increased coverage at certain segments of chromosomes 1, 2, 4, 6,7 and R. However these align with the locations of highly repetitive regions of the *Candida albicans* genome such as the rDNA and MRS regions. This indicated that there was no chromosome aneuploidy in this strain.

The *ΔΔtlo* strain (cassette recycled) genotype was confirmed by collaborator Prof. Matthew Anderson's lab at the University of Ohio who used Illumina technology to sequence the strain and confirm all of the *TLO* genes were disrupted (personal communication). They also confirmed that there was no aneuploidy in this strain. This investigation did however, uncover that the AHY940 WT strain did harbour an extra copy of chromosome 5.

4.3.5 Chromosome structure of strains

CHEF gels were carried out in the lab of Dr Alessia Buscaino at the University of Kent. Results of these gels can be seen in Figure 4.17. The chromosomal structure of AHY940 here again displays an extra copy of chromosome 5. This trisomy is resolved in the *ΔΔtlo* mutants, however these two strains do have different chromosome structures to each other. This does not appear to have had any effect on phenotype, as both strains performed in an almost identical manner in all phenotypic assays described in this chapter. The chromosome structure of the mutants is currently under investigation by Dr Samuel Vega Estevez at the University of Kent by *de novo* assembly of the Nanopore reads described here.

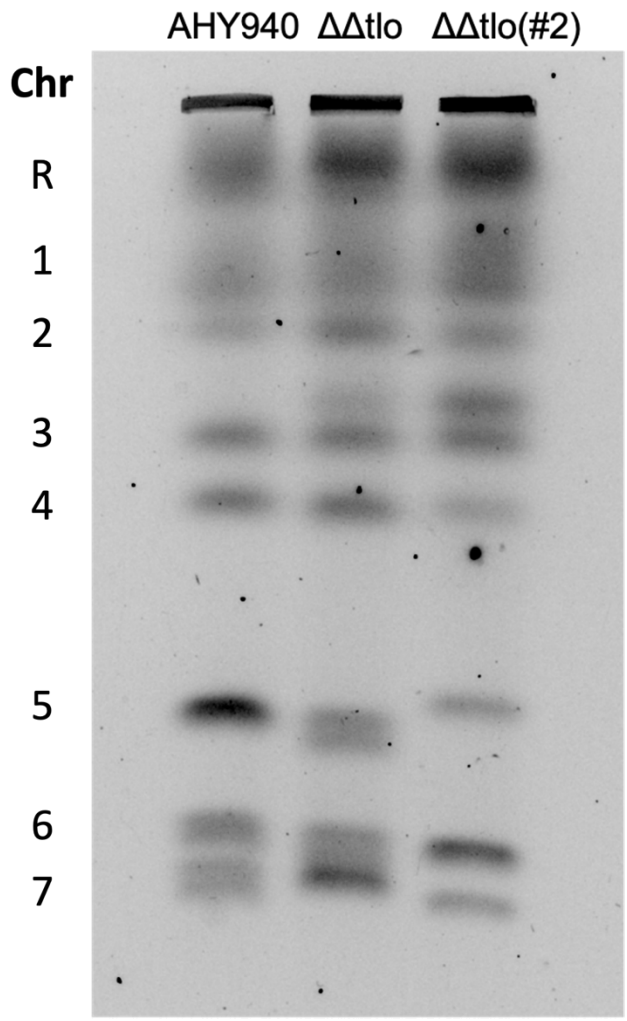


Figure 4.17 CHEF gel electrophoresis of *ΔΔtlo* strains and AHY940

CHEF gel electrophoresis of strains AHY940, *ΔΔtlo* and *ΔΔtlo* (#2). Chromosomes separate based on size, largest at the top. Corresponding chromosome numbers are indicated on the left hand side. All three strains show different chromosome structures. The dark band in the AHY940 lane indicates trisomy of chromosome 5.

4.4 Discussion

Deletion of the *TLO* gene family from the *C. albicans* strain AHY940 via a CRISPR Cas9 mutagenesis proved to be a highly specific and labour efficient method for deleting an entire gene family from this *Candida* background compared to traditional mutagenesis techniques. The ability to target the full family by targeting the Cas9 enzyme to a genetic sequence conserved in all members of the family. The defining feature of this gene family, the *MED2* domain, proved effective in creating a *ΔΔtlo* strain, which was reproducible in two separate transformation events. In some cases, incomplete knockouts were generated. These strains maintained some of their *TLO* genes, while others were truncated. Further investigation into this phenomenon would determine if these *TLO* sequences contained mutations that meant they could not be targeted with the CRIPSR guide. These strains may however be useful for studying the effects of a depleted *TLO* repertoire, with some *TLO* genes intact at their native loci. The generation of a second *ΔΔtlo* strain, *ΔΔtlo* (#2), from a separate mutagenesis event showed this method was reproducible, and the inclusion of this strain in phenotypic analysis confirmed that phenotypes observed were a result of *TLO* deletions rather than any secondary mutations.

Whole genome sequencing of the primary *ΔΔtlo* strain (with the cassette recycled) and the *ΔΔtlo* (#2) strain (with cassette in place) was performed, and the reads were aligned to the SC5314 reference genome. This showed that all of the *TLO* genes were disrupted successfully in both strains, Appx. 3 Figures 1 and 3. The aligned sequence was also used to look at depth of coverage across the entire genome in order to pinpoint any large duplicated or deleted regions, Appx. 3 Figures 2 and 4. This analysis highlighted only the MRS and rDNA regions as regions with exceptionally high read depth, but this is due to the highly repetitive nature of these genomic regions. Other than these regions, all other parts of each chromosome appeared to have even coverage, and the coverage levels across the chromosomes were consistent with each other, and the calculated average coverage. This indicated that there was no loss or expansion of regions of certain chromosome, or loss or duplication of chromosomes themselves. Analysis performed by the Anderson group at the University of Ohio confirmed that all the *TLOs* were truncated in the *ΔΔtlo* mutant, but also discovered that the parent strain, AHY940, harboured a trisomy of chromosome 5, however, this was resolved in the *ΔΔtlo* strain. The CHEF gels performed by the Buscaino group at the University of Kent to check the chromosome structures of the strains indicated that all three strains, AHY940, *ΔΔtlo*

(#1) and *ΔΔtlo* (#2) each had a different chromosome structure, Figure 4.17. The exact structure of these chromosomal rearrangements is currently under investigation. The cause of these rearrangements is unknown, but perhaps the trauma of such massive gene deletion via CRISPR caused genomic rearrangement. However, the fact that the two different *ΔΔtlo* mutants, with different chromosomal makeups, exhibited similar phenotypes in all phenotypic assays performed indicate that phenotypes seen are most likely due to the deletion of the *TLO* genes themselves rather than any chromosomal rearrangement that may have been induced by the mutagenesis. Complementation of these mutants with individual *TLO* genes, which was carried out partly to determine if the phenotypes described here are directly cause by loss of *TLO*s, is described in subsequent chapters.

A *ΔΔmed3* mutant derivative of SC5314 was included in the phenotypic analysis. The Med3 Mediator subunit is required to allow Med2 homologues (i.e. the Tlo proteins in *C. albicans*) to incorporate into the Mediator complex (Zhang *et al.*, 2012). It is thought that the incorporation of Tlo proteins into Mediator serves in some way to target the Mediator-RNA Polymerase II complex to DNA for transcription. In *S. cerevisiae*, the Mediator tail is required for transcription of TATA-containing, SAGA-dependent genes, where the tail subunits interact with DNA bound activators, directly or indirectly, to recruit RNA Polymerase II (Ansari *et al.*, 2012). It has also been suggested that free Tlo protein in the cell may be able to enact transcriptional functions independent of the Mediator complex due to the TAD contained in the protein (Liu and Myers, 2015). If correct, this hypothesis suggests that a *ΔΔmed3* strain is unable to incorporate Tlo proteins into the Mediator complex, resulting in free Tlo proteins in the cells that may be enacting their own Mediator independent functions. The inclusion of this *ΔΔmed3* mutant in phenotypic analysis allows differences in the *ΔΔtlo* and *ΔΔmed3* strains to be highlighted, where differences could possibly be a result of this free Tlo action, or the necessity for Tlo incorporation into the Mediator. Med3 itself also has functions independent to Tlo protein incorporation and Med3 is required for the maintenance of the Med15 subunit in the Mediator tail, so the *ΔΔmed3* strain also lacks Med15 in its Mediator (Zhang *et al.*, 2012). In *C. albicans* Tlo proteins are unable to be recruited into the nucleus if Med3 and Med15 are not present (Liu and Myers, 2015), therefore phenotypes seen in the *ΔΔmed3* strain may be due to the lack of Tlo in the nucleus also.

Phenotypic analysis of the mutants highlighted that the loss of Tlos generated a much less fit and less virulent strain. The results of these phenotypic tests have been compiled

into a heat map, which can be found in Figure 4.18. The *ΔΔtlo* mutants both grow as pseudohyphae in nutrient rich conditions, compared to the yeast form growth of the wild type strain in these conditions, where the WT strain grows in the yeast morphology, Figure 4.6. This indicates that loss of *TLO* genes in *C. albicans* has detrimental effects on cell growth and cell division. A *C. dubliniensis Δtlo1/Δtlo2* mutant strain also displays a pseudohyphal morphology compared to the WT *C. dubliniensis* strain growing in the yeast morphology (Haran *et al.*, 2014). This also indicates that the *TLO*s are necessary for the normal growth of these species. The observation that the *ΔΔmed3* strain also grows pseudohyphally could indicate that the incorporation of Tlo proteins into the Mediator complex, or the presence of Tlo in the nucleus, is essential for normal yeast growth in nutrient rich media.

Distinction between yeast cells, hyphae and pseudohyphae are described by Sudbery, Gow and Berman in (Sudbery *et al.*, 2004) and these criteria are used to define the morphological states of the strains. Pseudohyphae are thought to be defective in cell separation, where the daughter cells are produced, but after septum formation, the cells do not separate from the mother and continue to grow while still connected. Pseudohyphal cells are defined by constrictions along the budding junction between cells. It is still unclear if the pseudohyphal morphology is its own unique terminal cell morphology, or if it is an intermediate state between yeast and hyphae (Noble *et al.*, 2017). There are many genes involved in the yeast to hyphal transition, that, when deleted, result in constitutive pseudohyphal growth, such as such as; *TUP1,* a transcriptional co-repressor that represses transcription of hyphal specific genes*; NRG1,* another repressor of hyphal growth; *FKH2,* a transcription factor involved in regulation of morphogenesis; *GIN4* and *HSL1,* which both encode Nim1 kinases (Sudbery *et al.*, 2004, Noble *et al.*, 2017). However, as the *TLO* gene family is thought to play a role in the transcriptional regulation of many pathways, and that there are many different pathways that feed into the yeast-hyphal transition. More in-depth analysis into the transcriptome of the strains is required to determine what role the *TLOs* play in this complex network.

Attempts to induce true hyphal growth in the *ΔΔtlo* mutants was not successful, Figure 4.7. While the pseudohyphae did elongate, no true hyphae were formed. While true hyphae are long, narrow and have uniform parallel sides, pseudohyphal cells vary in

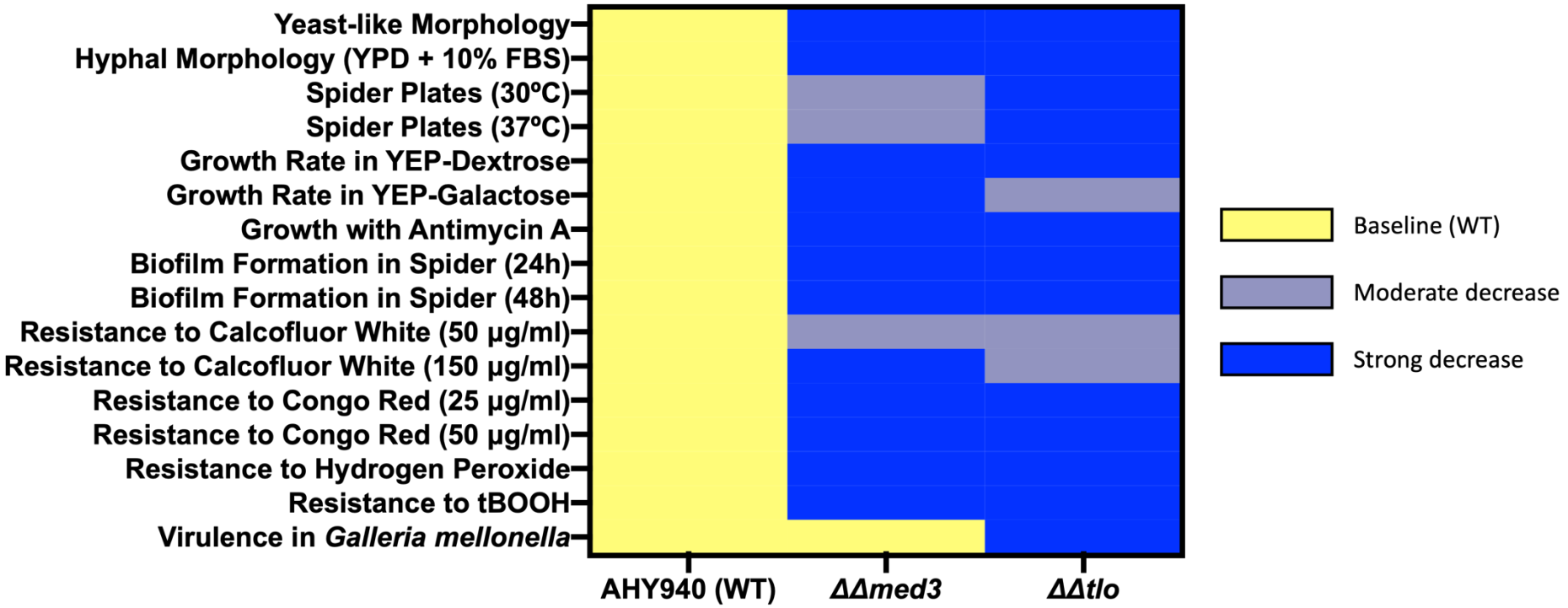


Figure 4.18 Heatmap summarising results of phenotypic analysis of *ΔΔtlo C. albicans* strains

Phenotypes observed in this work were compared to that of the WT strain AHY940 with severity of difference rated by colour. The scale is displayed on the right hand side. There is only one *ΔΔtlo* column as the results from both *ΔΔtlo* strains analysed displayed the same phenotypes in all cases tested.

size and shape, and they do not have uniform hyphal sides. Pseudohyphal elongations have indentations at septal junctions, while truly hyphal cells remain parallel and uniform along the entire length of the filament (Sudbery *et al.*, 2004).

It is possible for a constitutively pseudohyphal strain to transition to true hyphal growth, as is the case for an *ΔΔnrg1* mutant strain of *C. albicans*, which grows pseudohyphally in the absence of stimulants but can be induced to grow as true hyphae by introduction of serum, or by change in pH (Murad *et al.*, 2001). This is not the case for the *ΔΔtlo* strains. Attempts to induce true hyphal growth with YPD supplemented with FCS, were unable to induce the transition to true hyphal growth. Even incubation in water supplemented with FCS, which is a stronger inducer of the yeast to hypha transition, failed to induce truly hyphal growth. It appears that *TLO*s play an essential role in the yeast to hyphal switch induced by FCS in *C. albicans*, and that deletion of the *TLO*s results in a cell that cannot form true hyphae under these conditions.

The *ΔΔmed3* mutant strain could not form true hyphae in FCS inducing experiments either, but a difference between the *ΔΔmed3* and *ΔΔtlo* strains became apparent when attempting to induce hyphal growth on solid Spider agar, Figure 4.8. While the *ΔΔtlo* strains appeared smooth on this medium after incubation at 30 °C for 5 days, the *ΔΔmed3* strain displayed a wrinkled morphology that was more similar to the WT. A similar trend was seen after incubation on spider at 37 °C for 5 days, however the differences here were less clear cut, as the *ΔΔtlo* did display some wrinkling, but not to the extent of the *ΔΔmed3* mutant, which in turn was not as wrinkled as the WT. These results indicate that the presence of Tlo proteins in the *ΔΔmed3* mutant strain may play some role in the development of hyphae on spider agar, and that this phenotype may not completely rely on the incorporation of the Tlo proteins into the Mediator complex.

The ability to form chlamydospores was inhibited in both the *ΔΔtlo* and *ΔΔmed3* strains, Figure 4.9. Chlamydospores are formed in low nutrient conditions in the absence of glucose. While they are dormant cells, they are not typical spores, as their formation does not provide guaranteed long-term survival (Sonneborn *et al.*, 1999).

In the screening of a mutant library of homozygous insertion mutant *C. albicans* strains, Nobile *et al.* 2003 found that mutating genes *ISW2*, *SCH9* or *SUV3* completely abolished chlamydospore formation, and mutation of *RIM13*, *RIM101* or *MDS3* could delay chlamydospore formation. Their library, however, contained only 217 different gene

mutants and did not contain any *TLO* mutants. The only Mediator subunit included was *MED8* and this had no effect on chlamydospore formation (Nobile *et al.*, 2003).

In this work, the fact that chlamydospore formation is inhibited in both strains indicates that this phenotype may be dependent on integration of Tlo proteins into the Mediator complex via Med3. In the *Δnrg1* strain that is described by Murad *et al.* in 2001 their pseudohyphal *Δnrg1* mutant strain was able to form chlamydospores, indicating that the pseudohyphal morphology is not necessarily linked to the inability to form chlamydospores. However, a *Δtup1* mutant, which was also pseudohyphal, displayed delayed chlamydospore formation (Murad *et al.*, 2001). Neither of these genes were included in the Nobile 2003 screen (Nobile *et al.*, 2003).

The *ΔΔtlo* strains displayed a slower growth rate than that of the WT strain in YEPD, Figure 4.10. This suggests that the ability to grow in carbon rich, glucose rich media relies on the presence of *TLO*s. In *C. dubliniensis*, a *Δtlo1/Δtlo2* mutant strain also grew much more slowly in than its WT parent (Haran *et al.*, 2014), highlighting the necessity for *TLO*s for normal growth in nutrient rich media. The *ΔΔmed3* strain grew as slowly as the *ΔΔtlo* strains, indicating that this phenotype may require Tlo integration into Mediator, or Tlo localisation in the nucleus.

While again the *ΔΔtlo* strains grew significantly more slowly than the WT in YEP-Galactose, Figure 4.10, there was a difference in the growth rates between the *ΔΔmed3* mutant and the *ΔΔtlo* strains, where the *ΔΔmed3* grew more slowly than the *ΔΔtlo* mutants (*ΔΔmed3* also grew significantly slower than the WT). This difference could be a result of the loss of Med3 specifically, as it is thought that Med3 may have other functions separate to the role of integrating Tlos into mediator. Med3 itself does not have a TAD, but it has been shown by Liu *et al.* 2015, that the N-terminal region of Med3 has some transcriptional activation potential (Liu and Myers, 2015). This possibly explains the higher growth rate in the *ΔΔtlo* strains with Med3 intact, but the main increased doubling time phenotype is presumably caused by the lack of Tlo proteins incorporating into the Mediator.

While the *ΔΔtlo* was able to grow relatively unimpeded on solid YEPD agar at the varying temperatures tested, 22, 30, 37 and 40 °C, it did not grow to the same level as the WT, Figure 4.11. As seen with the exponential growth rate, it appears that *TLO*s are required for the normal growth of *C. albicans* cells. There was a morphological difference

apparent between the *ΔΔtlo* strains and the *ΔΔmed3* strain, which grew to a similar extent as the *ΔΔtlo* strain. This was most apparent at 30 °C, where the *ΔΔmed3* appeared as wrinkled colonies, while all the other strains appeared smooth. The *ΔΔnrg1* strain described in work by (Murad *et al.*, 2001) which was pseudohyphal in morphology in YEPD and formed wrinkled colonies on solid YEPD agar, as did a *ΔΔtup1* mutant (Braun and Johnson, 1997). While all of these strains were pseudohyphal on a cellular level, it appears that the colony morphology varied, and that a pseudohyphal/filamentous cellular morphology does not always result in a wrinkled colony.

The *ΔΔtlo* strains were both less able to grow than the WT on YEP-Galactose containing Antimycin A, a respiration inhibitor, Figure 4.12. Antimycin A inhibits part of Complex III in the respiration pathway, and cells cannot undergo traditional respiration. It is necessary to perform Antimycin A sensitivity assays in media where galactose is the carbon source as *C. albicans* has alternative pathways for taking up other hexose sugars such as glucose, whereas the uptake of galactose requires specific transporters (Askew *et al.*, 2009). The phenomenon of yeasts being unable to grow anaerobically or in the absence of respiration on certain carbon sources (galactose, raffinose, maltose) is known as the Kluyver effect, named for yeast *Kluvyeromyces lactis* (Goffrini *et al.*, 2002). Cells grown in the presence of Antimycin A have been shown to shift to a pathway that uses an alternative oxidase, made up of *AOX1a* and *AOX1b* subunits, which is much less efficient than respiration, and does not contribute to ATP synthesis (Chabrier-Rosello *et al.*, 2010). The ability of the WT strain to grow better than the mutant strains on Antimycin A, while still appearing somewhat inhibited due respiration being inhibited, could be explained by the induction of this AOX pathway in the WT. If this pathway is not as effectively induced in the mutant strains this could explain the much more hampered growth seen in the *ΔΔtlo* strains, and the *ΔΔmed3* strains.

When the comparative ability of the mutants to form biofilm was assessed, the *ΔΔtlo* mutants were much less able to form biofilm than the WT, Figure 4.13, indicating that the *TLOs* also play a role in biofilm formation pathways in the cell. Hyphal differentiation is believed to be essential in the biofilm formation process, as well as production of an extracellular matrix (Chandra *et al.*, 2001). The defects the *ΔΔtlo* strains display with regards to hyphal development and the fact that they appear to be constitutively pseudohyphal could explain their defective biofilm formation. There was no difference between the level of biofilm formed by the *ΔΔtlo* and the *ΔΔmed3*, indicating that the constitutive pseudohyphal morphology could be to blame. In biofilm assays described

in Chapter 3, it was seen that a *Δtlo1/Δtlo2* mutant *C. dubliniensis* strain was less able to form biofilm in the same conditions tested here. This analysis differed to that published in (Haran *et al.*, 2014) but could be due to differencing in assay design. This strain also displayed a pseudohyphal morphology, indicating that the presence of *TLO* genes is important in the biofilm formation process, be it though their role in morphological regulation, or through roles in other, more biofilm specific, pathways.

When the strains were tested on their ability to resist cell wall stress, the *ΔΔtlo* mutants were more sensitive to stress induced by Congo Red and Calcofluor White (CFW) than the WT strain, Figure 4.14. This indicates that loss of *TLOs* is detrimental to the cell's ability to resist cell wall perturbing compounds, or that the integrity of the *ΔΔtlo* mutant cell wall is compromised. In *C. dubliniensis,* deletion of *CdTLO1* and *CdTLO2* gave rise to a strain that was much more sensitive to CFW than the WT *C. dubliniensis* parent (Haran *et al.*, 2014). Congo Red targets the chitin and β-glucan in the cell wall, Calcofluor White also targets the chitin via the 1-4 β and 1-3 β polysaccharides bound to it (Ram and Klis, 2006). It could be speculated that increased sensitivity of these mutant strains is down to their pseudohyphal morphology and differences in cell wall structure. In the case of Congo Red, there was no difference between the *ΔΔtlo* and *ΔΔmed3* strains, but in the case of CFW, it appeared that at higher concentrations (150 μg/ml) the *ΔΔtlo* strains were more resistant than *ΔΔmed3* mutant. The possibility that pseudohyphal morphology is linked to cell wall stress sensitivity stands, as all pseudohyphal strains tested were much more sensitive than the WT strain. However, the differences in resistance to CFW between the *ΔΔtlo* strains and the *ΔΔmed3* strain could indicate that Med3 plays a role in CFW resistance, and its loss compounds the already increased sensitivity.

The *ΔΔtlo* mutants were much more sensitive to oxidative stress than the WT strain, Figure 4.15. The *ΔΔmed3* strain was similar to the *ΔΔtlo* mutants in sensitivity. Oxidative stress resistance is important for *C. albicans* in terms of resisting the host immune response. When neutrophils engulf *C. albicans* cells, reactive oxygen species (ROS) are produced to kill the invading pathogens. In response to oxidative stress, *C. albicans* upregulates the expression of many antioxidants; such as catalase (*CAT1*); glutathione peroxidases (*GPX*) and superoxide dismutases (*SOD*), as well as products which work to repair proteins damaged by oxidative stress (e.g. *GSH1, TTR1, TSA* and*, TRX1*). Strains that cannot resist oxidative stress induced by the host are much less able to cause infection and are much less virulent (Enjalbert *et al.*, 2007, Dantas *et al.*, 2015). The loss

of the *TLO* genes in the *ΔΔtlo* mutants appears result in increased susceptibility to oxidative stress, reinforcing that the loss of *TLOs* generates a less virulent strain of *C. albicans.* Similar phenotypes were observed in a *Δtlo1/Δtlo2* mutant strain of *C. dubliniensis*, where loss of the *TLO* family made the strain less able to resist oxidative stress (Haran *et al.*, 2014).

Deletion of the *TLO* repertoire from *C. albicans* gave rise to a strain that is much less virulent than WT in a *Galleria mellonella* wax moth larvae infection model, Figure 4.16. This indicates that the loss of *TLO* genes reduces virulence. A *Δtlo1/Δtlo2* strain of *C. dubliniensis* was also much less virulent than its WT counterpart in *G. mellonella* infection models, as seen in Chapter 3 of this thesis, and in (Flanagan *et al.*, 2018). Comparative transcriptomic analysis of *C. albicans* infecting either mice or *G. mellonella* showed an increase in *TLOα1* expression in the *Galleria* model, but this gene was downregulated in the mouse model. The researchers who performed this work suggest that the increase in expression of *TLOα1* may play a niche-specific role during infection of *G. mellonella* (Amorim-Vaz *et al.*, 2015). The expression of *CaTLOs* in the WT *C. dubliniensis* background, discussed in Chapter 3, indicates that increasing expression of *TLO*s increases virulence. It could be suggested that not all of this overexpressed Tlo protein is being incorporated into Mediator (this has not been proven via protein quantification methods) and that it is the increase in abundance of free Tlo that could be resulting in the increase in virulence in the *Galleria mellonella* model. Differences have been observed between the Galleria and the gold standard murine systemic infection models (Amorim-Vaz *et al.*, 2015), so it would be interesting to see if the *ΔΔtlo* mutants are also less virulent in the latter model.

These virulence assays also showed that the *ΔΔmed3* mutant exhibited greater virulence than the *TLO* mutants. This points toward this phenotype being a free Tlo phenotype, and that a cell with free Tlo protein, not associated with Mediator, can be just as virulent in a *G. mellonella* model as a WT strain, and indicates that Tlo proteins are performing another function elsewhere during the infection of *G. mellonella.*

To summarise, it has been demonstrated that the deletion of the entire *TLO* gene family in *C. albicans* is not lethal. However, phenotypic analysis of the mutants suggests that a wide range of morphological and biological functions are affected by the deletions, including a loss of fitness and virulence when compared to the WT. This indicates that the *TLO*s are likely to play a significant role in the routine function of the cell. The wide

variety of phenotypes affected by the deletion of *TLOs* highlights that they are involved in a wide array of cellular pathways. The comparison to the *ΔΔmed3* mutant strain, with its *TLO* repertoire still intact, highlighted that there may be some phenotypes, (e.g. morphology and growth rate) affected by Tlo in the Mediator or nucleus, and some that are affected by the possible presence of a pool of free Tlo elsewhere in the cell (e.g. filamentation on Spider agar and *G. mellonella* infection), and it appears that a complete Mediator complex is necessary for the optimal operation of the cell. The following chapters describe transcriptomic and proteomic experiments designed to try to investigate the role of the Tlo proteins in general and to investigate if different members of the *TLO* gene family encode proteins with specific functions.

Chapter 5

Reintroduction of *TLO* genes into a *ΔΔtlo* mutant background *Candida albicans* strain

5.1 Introduction

Previous chapters in this thesis have proposed that the expansion of the *TLO* gene family of *Candida* spp. can enhance virulence, and that complete loss of *TLO*s results in a less fit mutant strain. Expansion of the *C. dubliniensis TLO* repertoire with *CaTLO*s as described in Chapter 3, indicated that *TLO* genes from different clades may affect different phenotypes. This work also indicted that a higher level of expression of *TLOs* conferred a higher level of virulence in a *Galleria mellonella* infection model. The generation of a *ΔΔtlo* mutant strain of *C. albicans* described in Chapter 4 created a background to study the clade-dependent effects of different *TLO* genes in isolation and demonstrated that the presence of *TLO*s is required for the normal function of the cell.

5.1.1 Candida albicans TLO *genes and clades*

The expansion of the *TLO* gene family in *Candida albicans* is unique in eukaryotes, and while most other *Candida* spp. only have one *TLO* gene, and *C. dubliniensis* has two, *C. albicans* strains can have between 10-15 *TLO* gene family members. This expansion is the largest expansion of gene copy number between *C. albicans* and other members of the CTG clade (Hirakawa *et al.*, 2015).

Tlo proteins contain a Med2 domain at the N-terminal that allows for integration into the Mediator complex. The C-terminal domain of each Tlo is distinct, but phylogenetic analysis of the 14 *TLO* genes in SC5314 separates them into three distinct clades, α, β and γ, based on long terminal repeat (LTR) insertions in the C-terminal domains (Anderson *et al.*, 2012). Figure 5.1 shows a map of the *TLO* gene loci and clades for the SC5314 lab strain of *C. albicans* (A) and a diagram highlighting the genetic differences in the clades (B). The N-terminal Med2 domain itself does not possess a transcriptional activation domain, but it does facilitate Tlo localisation to the nucleus via integration into the Mediator complex or through a nuclear localisation sequence (Anderson *et al.*, 2012, Liu and Myers, 2015). The C-terminal domains do possess transcriptional activation domains, but no means of nuclear localisation, meaning that both ends of the protein need each other to properly enact their function (Liu and Myers, 2015). The *TLO* gene family is believed to have expanded from one ancestral *TLO* gene through subtelomeric recombination events. These recombination events are still occurring, and recombination and loss of heterozygosity can affect the copy number and clade

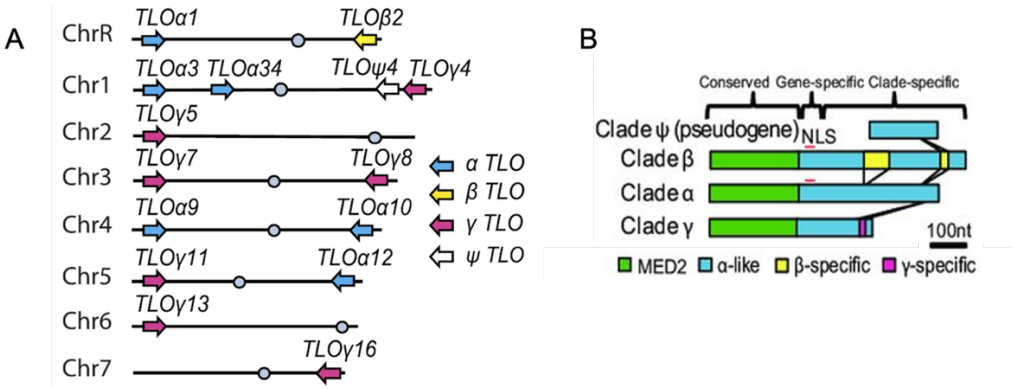


Figure 5.1 Locations and structure of *TLO* genes in *Candida albicans*

(A) A map of *TLO* genes across the *C. albicans* SC5314 genome. *TLO* genes are indicated by arrows, and centromeres are grey circles. The clade of each *TLO* gene is indicated by the colour of their arrow according to the legend in the diagram. This figure is taken from (Anderson *et al.*, 2012). (B) The *TLO* genes of *Candida albicans* can be split into three clades and a pseudogene based on their C-terminal domains. All *TLO*s (except the pseudogene) have a conserved Med2 domain (green) at the N-terminus. Following this, there is a gene specific region, which is highly variable, but has some similarity to the α clade (blue). The β clade contains specific insertions (yellow) distinguishing it from the *α* a clade. *γ* clade *TLOs* also differ from the other clades by insertions (pink). Black lines show where insertions are located in relation to the other clades. The *TLOψ4* pseudogene does not share the Med2 domain with other *TLO*s but has some homology to the C-terminal domain of the α clade TLOs. The predicted nuclear localisation sequence found in α and β clade TLOs is indicated by a red line. Adapted from (Anderson *et al.*, 2012).

representation of *TLO*s in *C. albicans* strains (Anderson *et al.*, 2015).

In the commonly used lab strain, SC5314, there are seven members of the γ clade, which are the shortest *TLO* genes with an average length of around 525 bp (Anderson *et al.*, 2015). The γ clade is the only clade of *TLO* gene that have been shown to undergo alternative splicing (see below). The α clade of *TLO*s is comprised of five members and measure between 675 and 750 bp. The β clade contains one single *TLO*, *TLOβ2*, and this gene is 822 bp in length. *TLO* genes within a clade can share up to 97% similarity, and between clades the *TLO* genes are around 82% identical (Anderson *et al.*, 2015).

All members of the γ clade have been predicted to undergo alternative splicing after transcription, and it has been proposed that each different mRNA isoform can be translated to a functional protein. The proteins encoded by the alternative isoforms can differ in size from each other by between 8 and 43 amino acids (Anderson *et al.*, 2012).

It has been suggested that the different clades of *TLO* display different patterns of cellular localisation, with the α and β clade *TLO*s containing a nuclear localisation sequence that directs the proteins to the nucleus. Spliced *γ* clade *TLO* transcripts encode proteins that primarily localise to the nucleus, however proteins from un-spliced γ clade *TLO* transcripts have been shown to localise to both the nucleus and the mitochondria. (Anderson *et al.*, 2012).

TLO genes encode proteins that are interchangeable subunits of the Mediator complex (Zhang *et al.*, 2012). Research by Anderson *et al.* in 2012 showed that the ratio of non-synonymous to synonymous evolutionary changes in the Med2 domain of the *TLO* genes shows a selection pressure towards conservation of the Med2 domain in the family (Anderson *et al.*, 2012). Expansion of the *TLO* genes into a family that shows diversity at a copy number and clade level while still maintaining a Med2 domain to allow the proteins to serve as Mediator subunits is thought to generate a pool of Mediator complexes that all contain different Tlo proteins in the Med2 position at the tail of the complex (Liu *et al.*, 2016). One of the aims of this thesis was to investigate whether having a pool of different Mediator complexes with different Tlo proteins might result in transcriptional plasticity of *C. albicans* resulting in the ability of this organism to adapt to different host niches and colonise and infect humans more effectively than other *Candida* species (Hirakawa *et al.*, 2015).

5.1.2 Free Tlo

In *C. albicans,* while the Tlo proteins do incorporate into the Mediator complex, there is also a large pool of unincorporated, 'free' Tlo. In fact, the majority of Tlo protein in *C. albicans* is in this 'free' state, with up to 10 times more Tlo existing outside of Mediator than within it (Zhang *et al.*, 2012).

In *C. dubliniensis*, Liu *et al.* found that CdTlo1 exists only as a part of the Mediator complex, and cannot be detected in an unassociated form, even when artificially overexpressed (Liu *et al.*, 2016). While under normal conditions *CdTLO2* is expressed around 50 times less than *CdTLO1* mRNA and exists at around 30 times less on the protein level, it is possible to artificially generate a pool of free CdTlo2 protein in the *C. dubliniensis* cell. Overexpression of *CdTLO2,* and subsequent generation of a pool of free CdTlo2, could allow the *C. dubliniensis* cell to filament to a similar extent as *C. albicans*, where previously filamentation of *C. dubliniensis* required much more stringent stimulation. These experiments were performed *in vitro* in nutrient rich media, and may not represent other conditions (Liu *et al.*, 2016). The phenotypic effect of generation of this free pool of Tlo protein in *C. dubliniensis* demonstrates the possible consequences of the free Tlo pool in *C. albicans* and shows that free Tlo could be enacting functions outside of association with Mediator.

Candida albicans α and β clade Tlos have been shown to contain a transcriptional activation domain (TADs) which can function independently to Mediator (Liu *et al.*, 2016). These TADs have also been observed in *C. dubliniensis* Tlos and the Med2 protein in *S. cerevisiae*. It has been suggested that the presence of TADs in free Tlo proteins, acting independently of Mediator, allows for them to compete for DNA binding sites on co-activators and co-repressors, thus enacting transcriptional changes on their own accord (Liu *et al.*, 2016). The large pool of free Tlo present in *C. albicans* may also be 'squelching' the response of the Mediator complex, where Tlo proteins sequester DNA binding sites on co-activators that Mediator is targeting via the tail module, and so blocking transcriptional activation (Zhang *et al.*, 2012).

5.1.3 Noisy expression of TLOs

The regulation of *TLO* gene expression has not yet been investigated in detail. However, the *TLO* promoters have been shown to contain a *GAL4* binding site, suggesting that their

expression may be, at least partly, under the control of the Gal4 transcription factor (Sullivan *et al.*, 2015).

As well as variation between populations, there is also a level of cell-to-cell variation in *C. albicans*. The expression has been described as 'noisy' and the expression pattern has been named telomere-adjacent gene expression noise (TAGEN) (Sullivan *et al.*, 2015). TAGEN is also evident in the expression of Sir2, which is an NAD+-dependent histone deacetylase that acts to modify chromatin, removing acetyl groups from histones in order to shut down transcription (Anderson *et al.*, 2014). This noisy expression influences the availability of Tlo to integrate into the Mediator complex and the pool of free Tlo present in the cell, which could likely have an effect on downstream transcription in the cell.

The location of the *TLO* genes might also play a role in their regulation, as being close to the telomere suggests that they are likely to be subjected to silencing or activation due to chromatin-mediated positional effects. Genes located at less heterochromatic regions of the genome are not under this additional level of transcriptional control. Telomere Positional Effect (TPE) is a phenomenon by which genes close to telomeres are silenced due to the congregation of specific chromatin complexes at the telomeres that extend into the chromosome. In *S. cerevisiae*, the Mediator complex interacts with the Sir2 protein to regulate this effect (Anderson *et al.*, 2014). As the Tlo proteins are subunits of the Mediator complex, it is possible that they could play a role in affecting their own expression. It was seen in work performed in *C. dubliniensis* that the CdTlo1 protein localised to the telomeres and more specifically, to ORFs within the telomeric regions (Haran *et al.*, 2014). It was also observed that genes that CdTlo1 was found to localise at were generally highly expressed in the WT, and downregulated in a *Δtlo1/Δtlo2* null mutant, however there was a range of expression levels. The CdTlo1 protein was also found to localise in the WT to genes that were not expressed, such as those related to starvation responses, the expression of which was increased in the absence of Tlo in the *Δtlo1/Δtlo2* null mutant. This suggests a dual role for CdTlo1 in activation and repression. We know that CdTlo1 is only present in a Mediator-associated form in *C. dubliniensis*, and research in *S. cerevisiae* has shown that the yeast Mediator can localise to both expressed and repressed genes, and is believed to adopt different conformations, via the Cdk8 module, to enact different regulatory functions (Haran *et al.*, 2014).

5.1.4 Aims of this work

The main aim of this chapter was to attempt to investigate if different Tlo proteins possess specific functions. This was achieved by generating a strain set in the *ΔΔtlo C. albicans* background where a *TLO* from each clade was reintroduced under a promoter driving a high or relatively low level of expression. Different strengths of promoter were used to determine if the level of expression of each *TLO* had any effect on phenotype. Once each strain was constructed, phenotypic tests was performed on the set of strains in order to determine which phenotypes, if any, were restored or augmented by the introduction of specific *TLO*s.

5.2 Materials and Methods

5.2.1 Plasmid cloning

Sequences for the plasmids and inserts used in this work can be found in Table 2.3, and the sequences for the oligonucleotide primers used can be found in Table 2.2. Schematic diagrams for the *TLO* reintegration cassettes can be found in Figure 5.2.

5.2.1.1 pSFS2a cloning (*ENO1* promoter)

Constructs were designed to reintroduce clade representative *TLO* genes back into the *ΔΔtlo* mutant background. *TLOα1, TLOβ2* and *TLOγ11* were selected as representatives of their respective clades, and these were reintroduced under the strong Enolase promoter (*ENO1p*). A plasmid derivative of pSFS2a (Reuss *et al.*, 2004) was designed for the integration of each *TLO* gene, where homology to the *TLOα34* locus flanks the *TLO* construct and *SAT1* flipper cassette. The *TLOα34* locus was selected as the site for reintroduction of the *TLO* genes due to its location in the middle of the chromosome, to provide stability from recombination and stable expression, compared to all the other *TLO* sites, which are located at the chromosome ends and subjected to silencing and high rates of recombination.

A 300bp segment of the *TLOα34* 3′ sequence was amplified from the wild type (AHY940) *C. albicans* genome, and restriction sites were introduced via the amplification primers. A *Sac*II digestion site was introduced at the 5′ end of this segment, and a *Sac*I site was introduced at the 3′ end. pBluescript was used as an intermediate vector for cloning, as direct integration of this fragment into pSFS2a proved difficult. The digested fragment was ligated to the pBluescript backbone digested with *Sac*I/*Sac*II. This ligation mix was transformed into DH5α *E. coli* cells as described in Section 2.3.9. Transformants were selected on L agar with 100 μg/ml ampicillin, 40 μg/ml X-Gal and 40 μg/ml IPTG. Transformants which grew on L ampicillin, but were white in colour due to LacZ disruption were selected. Plasmid extraction and digestion with *Sac*I/*Sac*II confirmed presence of the fragment, and a gel extraction was performed on the *TLOα34* fragment. The cleaned up, digested fragment was ligated to a *Sac*I/*Sac*II digested pSFS2a backbone using T4 DNA ligase, see Section 2.3.5, and this ligation mix was transformed into DH5α competent *E. coli* cells. Transformants were selected on L agar containing 50 μg/ml chloramphenicol.

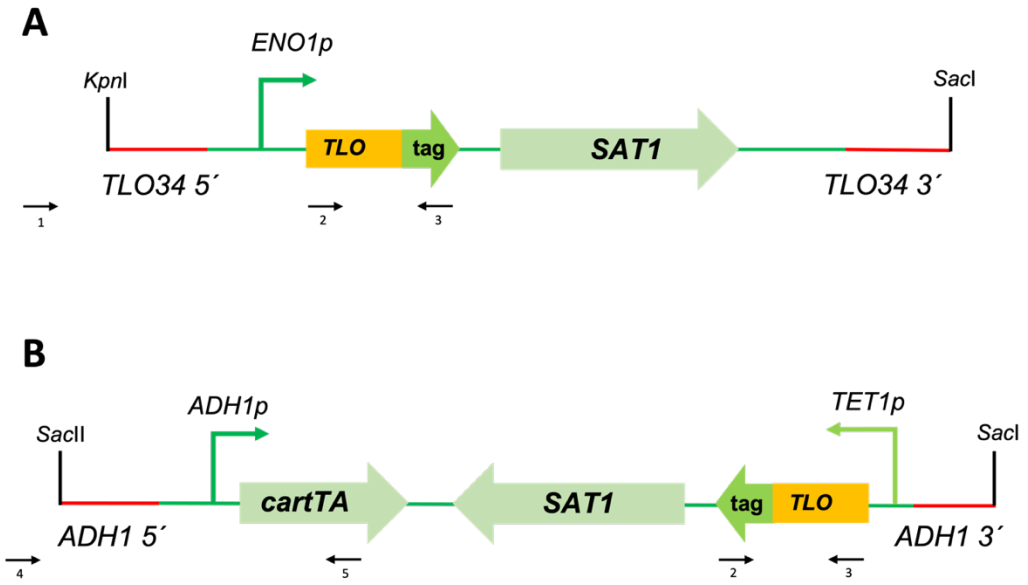


Figure 5.2 Cassettes used for reintroducing *CaTLO* genes into the ΔΔ*tlo C. albicans* mutant

(A) The *TLO* gene with a 3xHA tag under the *ENO1* promoter (high expression). This cassette integrates into the *TLO34* locus based on homology to this locus at the 5´ and 3´ ends. The *Kpn*I and *Sac*I enzymes were used to digest and liberate the cassette from the pJESS plasmid before transformation. The *SAT1* cassette contains a nourseothricin resistance gene and a maltose inducible flippase for recycling of the nourseothricin gene. (B) The *TLO* gene with 3xHA tag under the control of a tetracycline inducible promoter, *TET1*. This cassette integrates into the *ADH1* locus. The *Sac*I and *Sac*II restriction enzymes can be used to digest the pNIM1 plasmid that houses this cassette to free the cassette before transformation. Black numbered arrows indicate primer binding sites for confirmatory PCRs for integration of cassettes into correct sites. Primer names are as follows: 1) TLO34 locus F; 2) TLO (*Sal*I) F; 3) TLO tag (*Bgl*II) R; 4) ADH1 locus F; 5) carTA R.

Plasmid extraction and digestion confirmed the presence of the plasmid and the correct integration of the insert. This plasmid was sent to Genewiz for the synthesis and cloning in of the designed *TLO* cassettes (see Fig. 5.1.1). Using SerialCloner, sequences for the selected *TLO*s preceded by *ENO1* or the *TLOα1* promoter were designed, a 3x HA tag was added to the 3' end of the gene, and the *ENO1* terminator. A sequence of around 450 bp with homology to the 5′ region of the *TLOα34* locus was designed upstream of the promoter to enable integration of the cassette into the genome at the *TLOα34* locus.

Genewiz was unable to complete the set of plasmids under the control of the *TLOα1* promoter, and so a new strategy was devised to express the *TLO*s in the null mutant background using a doxycycline inducible promoter (see Section 5.2.1.2).

Plasmids were linearised with *Kpn*I and *Sac*I before transformation into the *C. albicans ΔΔtlo* background. Possible transformants were selected on YEPD agar with 200 μg/ml nourseothricin.

5.2.1.2 pNIM1 cloning (*TET1* promoter)

Oligonucleotide primers were designed that would amplify each tagged *TLO* gene from the respective pSFS2a plasmid, incorporating a *Bgl*II and a *Sal*I restriction site at the ends for cloning (see Table 2.2). Amplification was performed using routine PCR conditions and the size of the products was confirmed by agarose gel electrophoresis. The PCR products were then cleaned up and the concentration of nucleic acids was quantified.

The pNIM1 plasmid (Park and Morschhauser, 2005) was digested with *Bgl*II and *Sal*I, the fragments were separated by agarose gel electrophoresis. The largest fragment was separated from the others by gel purification. The concentration of this DNA was quantified.

Each PCR *TLO* product was ligated to the pNIM1 backbone using T4 DNA ligase according to the manufacturer's instructions. The ligation mix was transformed into *E. coli* DH5α cells and grown on L agar with 100 μg/ml ampicillin for selection. Ligated plasmids were confirmed by restriction digestion as well as by routine PCR using the *TLO* primers to check the plasmid contained the desired *TLO*.

These plasmids were linearised with *Kpn*I and *Sac*II before transformation into the *C. albicans ΔΔtlo* background. Possible transformants were selected on 200 μg/ml nourseothricin YEPD.

5.2.2 Confirmation of integration of cassettes

Strains expressing the *TLO* genes under the *ENO1* promoter were confirmed using two sets of PCR primers, the first using the *TLO* amplification primers, and the second using the reverse primer from this set and the TLO34 5′ F primer used in Section 5.2.1.1 to confirm reintegration into the *TLO34* locus.

Strains expressing the *TLO* genes under the *TET1* promoter were also confirmed by routine PCR with the *TLO* amplification primers, as above, as well as *ADH1* F and carTA R primers to confirm reintegration into the *ADH1* locus.

5.2.3 Gene and protein expression analysis

qRT-PCR was carried out as described in 2.4 to confirm transcription levels from the reintegrated genes.

The expression analysis indicated that there was leaky expression from the doxycycline inducible promoter in the absence of doxycycline sufficient to see mRNA transcription and protein expression. It was decided that for phenotypic analysis and comparison of a high and a low level of expression of *TLO*s, this leaky level of expression was sufficient to constitute a low level. The *TLO*s under the control of the *ENO1* promoter represent high levels of expression.

Western Blot analysis was carried out as in Section 2.5 to confirm translation of protein from the transformed constructs.

5.2.4 Phenotypic analysis

Phenotypic analysis was carried out on these strains as described in Section 2.6.

5.3 Results

5.3.1 Confirmation of reintegration of TLO *constructs in* ΔΔtlo

After transformation, confirmation of the integration of the *TLO* cassettes into the *ΔΔtlo* mutant was performed by routine PCR. Presence of the tagged *TLO* gene was confirmed by PCR with a general *TLO* forward oligonucleotide primer and the 3xHA tag specific reverse primer. Products of this PCR can be seen on an agarose gel in Figure 5.3 (A), where only those strains containing a 3xHA tagged *TLO* construct produced a product.

PCR tests were also performed to confirm the locations of the integrated cassettes The *TLO* constructs under the *ENO1* promoter were designed to insert at the *TLOα34* locus, and integration was confirmed using a forward oligonucleotide primer adhering upstream of this locus, and the reverse primer specific to the 3xHA tag. The *TLO* genes under the *TET1* promoter were inserted into the *ADH1* locus, the integration of the cassette at the *ADH1* locus was confirmed using a *ADH1* locus binding forward primer and a cassette specific reverse primer. The products of these reactions can be seen in Figure 5.3 (B) and (C).

5.3.2 qRT-PCR analysis confirming mRNA expression from TLO *inserts*

Expression of mRNA from the reintegrated *TLO* genes was quantified using qRT-PCR. Expression of the desired *TLOs* was seen in the respective strains, with a noticeable difference in expression in the levels from the different promoters, as seen in Figure 5.4. Those under the *ENO1* promoter were expressed at much higher levels than those under the *TET1* promoter.

The expression from the *TET1* promoter is doxycycline inducible, however there is some degree of 'leaky' expression from this promoter in the absence of doxycycline. The expression level displayed in Figure 5.4 is in the absence of doxycycline. For the purposes of this work, the leaky expression from the *TET1* promoter in the absence of doxycycline was deemed to be sufficient to show the difference in effect of high expression and low expression of the *TLO* genes in phenotypic analysis.

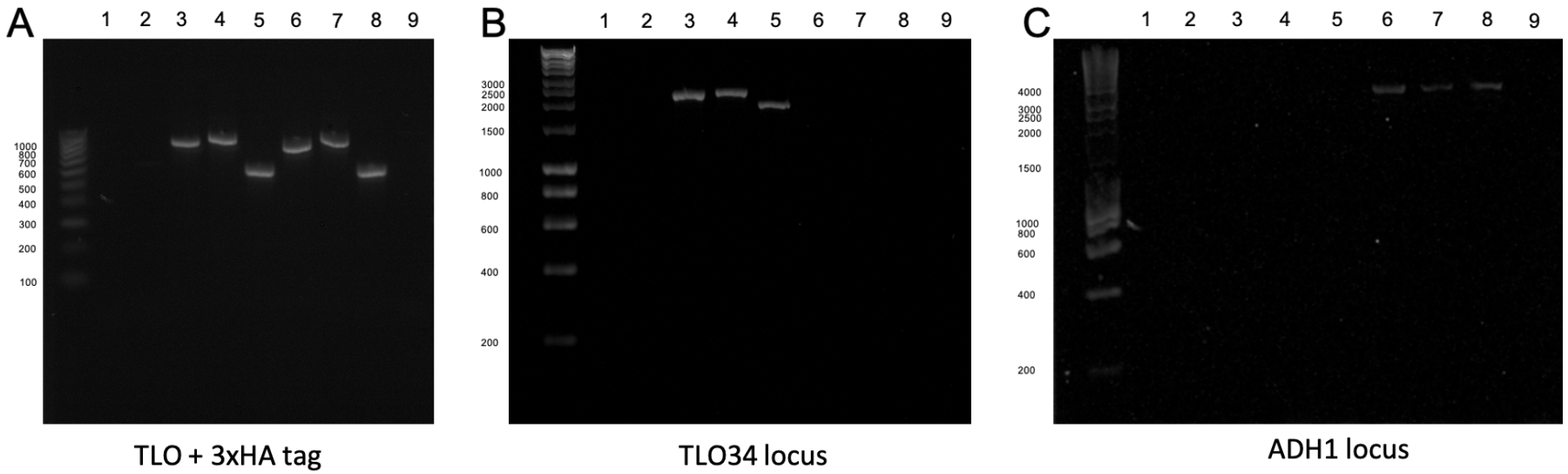
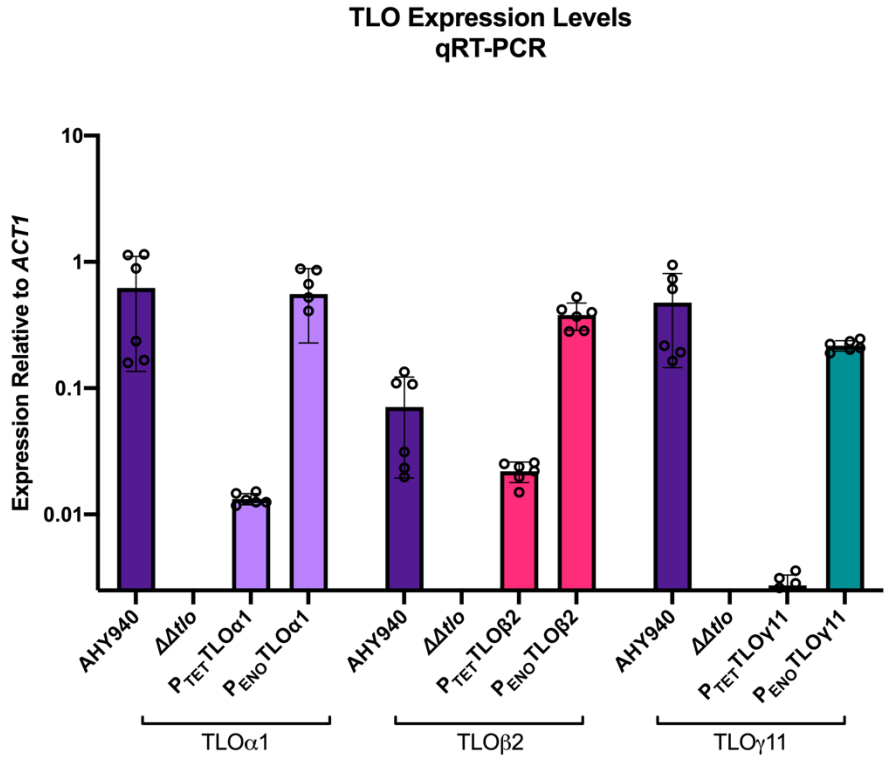


Figure 5.3 PCR confirmation of reintroduction of *TLO* constructs

Routine PCR was performed to confirm successful integration of the *TLO* constructs into the *ΔΔtlo* background. Lanes in all gels are as follows: 1) AHY940 WT; 2) *ΔΔtlo*; 3) P_{ENO1}*TLOα1*; 4) P_{ENO1}*TLOβ2*; 5) P_{ENO1}*TLOγ11*; 6) P_{TET1}*TLOα1*; 7) P_{TET1}*TLOβ2*; 8) P_{TET1}*TLOγ11*; 9) negative control (no DNA template). (A) PCR with oligonucleotide primers "TLO (Sal*I*) F" and "TLO tag (Bgl*II*) R" targeting the *TLO* construct via a general *TLO* forward primer and a reverse primer specific to the 3xHA tag at the end of the reintegrated *TLO*. Predicted sizes: *TLOα1* = 844 bp; *TLOβ2* = 913 bp; *TLOγ11* = 601 bp. (B) PCR to detect reintegration of the P_{ENO1}*TLOs* at the *TLOα34* locus using primers "TLO34 locus F" and "TLO tag (Bgl*II*) R". Predicted sizes: P_{ENO1}*TLOα1* = 2110 bp; P_{ENO1}*TLOβ2* = 2180 bp; P_{ENO1}*TLOγ11* = 1870 bp. (C) PCR to detect reintegration of P_{TET1}*TLOs* at the *ADH1* locus using primers "ADH1 locus F" and "carTA R". Predicted sizes: Positive = 4451 bp.

A



B

	AHY940	*ΔΔtlo*	P_{TET}TLOα1	P_{ENO}TLOα1	P_{TET}TLOβ2	P_{ENO}TLOβ2	P_{TET}TLOγ11	P_{ENO}TLOγ11
TLOα1 Mean	0.622	0.0006	0.013	0.557	-	-	-	-
TLOα1 S.D.	0.486	0.0004	0.001	0.329	-	-	-	-
TLOβ2 Mean	0.071	0.0016	-	-	0.022	0.380	-	-
TLOβ2 S.D.	0.052	0.0007	-	-	0.004	0.093	-	-
TLOγ11 Mean	0.477	0.0002	-	-	-	-	0.003	0.217
TLOγ11 S.D.	0.331	0.0001	-	-	-	-	0.0006	0.021

Figure 5.4 qRT-PCR analysis of mRNA expression of reintroduced *TLOs*

mRNA expression levels of each reintegrated *TLO* were measured in relation to an endogenous control gene (*ACT1*). This demonstrates the difference in expression level between the *TET1* and *ENO1* promoters. This analysis was performed in the absence of any doxycycline, which shows that leaky expression from the *TET1* promoter drives sufficient *TLO* mRNA expression for the purposes of this work. (A) Graphical representation of qRT-PCR results displaying the expression of each TLO gene in the relevant strains. The y-axis is in $\log_{10}$ increments. (B) The mean and standard deviation of the TLO genes in the relevant strains is given.

5.3.4 Western Blot analysis confirming protein expression from TLO *inserts*

Protein expression from the integrated *TLO* cassettes was investigated by Western Blot analysis targeting the 3xHA tag introduced at the 3′ end of the *TLO* genes in the constructs. These experiments were performed in various media; YEPD, YNB and Spider medium, with all experimental set ups giving concordant results. Figure 5.5 displays the results of a Western Blot performed in liquid Spider medium can be seen. Two strains of *C. dubliniensis* Wü284 expressing 3xHA tagged *CdTLO1* and *CdTLO2* respectively were included in this analysis as a positive control for the anti-HA antibody. This analysis included an RNA Polymerase II antibody probe a loading control, and this antibody was able to detect the RNA Polymerase II protein in all lanes.

The anti-HA antibody probe was able to detect the tag in the positive control lanes and the *TLOα1* and *TLOβ2* reintroduction strains. There was a detectable difference in the intensity of the bands in these lanes, with the strains expressing the *TLO* under the *ENO1* promoter appearing more intense. There was no HA tag detected in either the AHY940 lane or the *ΔΔtlo* lane. There was no HA tag detected in the *TLOγ11* reintroduction strain lanes, despite *TLOγ11* RNA being expressed, as indicated by qRT-PCR (see Fig.5.4). Strains were grown in a variety of media (e.g. YEPD, YNB) but no Tloγ11 protein was detected (data not shown).

4.3.5 Phenotypic analysis

In order to investigate if the three *TLO* genes expressed in the *ΔΔtlo* mutant background exhibited different functions, the phenotypes of the reintegrated constructs were examined using a wide range of assays similar to those used in the analysis of the *ΔΔtlo* mutant.

5.3.5.1 Colony and cellular morphology

The morphological appearance of the strains after 48 h of growth at 37 °C in YEPD liquid and solid medium was examined. All strains appeared as smooth colonies on solid YEPD agar, except for the *TLOβ2* reintroduction strains. The *TLOβ2* reintroduction strains appeared as wrinkled colonies, with a slight difference between the strain expressing high levels of *TLOβ2* versus the strain expressing low levels, where the higher level of expression yielded an objectively more wrinkled phenotype. Morphologies can be seen

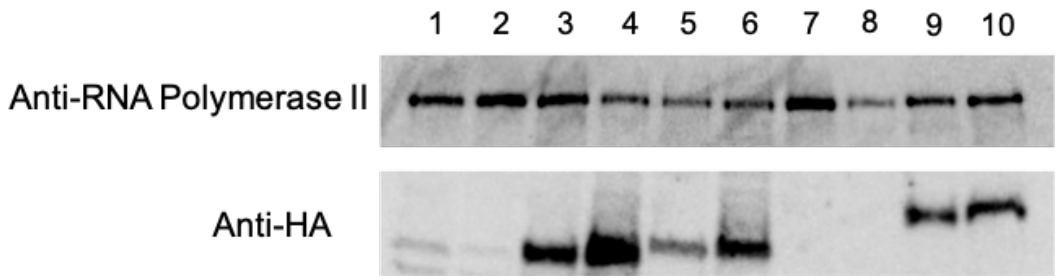


Figure 5.5 Western Blot analysis to detect Tlo protein expression in *TLO* reintroduction strains

Crude protein extracts from *Candida* strains containing the *TLO* reintegration constructs grown in Spider liquid broth were used for Western Blot analysis with an anti-HA antibody, which would bind to the 3x-HA tag on each construct. An image of the blot can be seen above. The membrane was stripped and re-probed with an anti-RNA Polymerase II antibody to act as a loading control. Proteins were extracted from cell suspensions that were standardised to an OD_{600} of 2 before extraction, therefore the intensity of the bands should be reflective of the level of expressed protein. Lanes as follows; 1)| AHY940 (WT) 2) *ΔΔtlo*; 3) P_{TET1}*TLOα1*; 4) P_{ENO1}*TLOα1*; 5) P_{TET1}*TLOβ2;* 6) P_{ENO1}*TLOβ2*; 7) P_{TET1}*TLOγ11*; 8) P_{ENO1}*TLOγ11*; 9) *C. dubliniensis CdTLO1-HA*; 10) *C. dubliniensis CdTLO2-HA*.

in Figure 5.6. Cells were viewed under a microscope after overnight growth at 37 °C with 200 rpm shaking in liquid YEPD medium in order to examine cellular morphology. The WT and *ΔΔtlo* mutant strains differed in that the WT strain grew in a yeast morphology, while the *ΔΔtlo* strain grew as pseudohyphae. Expression of the *TLOα1* gene, under either a strong or weak promoter, caused the cells to grow in the yeast morphology. Expression of *TLOγ11* under either promoter had no effect on cellular morphology at 37 °C, where the cells grow pseudohyphally, like the *ΔΔtlo* mutant. Expression of *TLOβ2* under either promoter had a dramatic effect on cellular morphology. The cells did not grow pseudohyphally like the *ΔΔtlo* strain, or fully as yeast cells like the WT, instead they appear to grow predominantly as yeast cells, with a subpopulation of true hyphae also present. As with the colony morphology, the strain expressing *TLOβ2* at a higher level appeared more hyphal than the strain expressing *TLOβ2* at the lower level.

5.3.5.2 Hyphal induction

To test the ability of the mutant and reconstituted strains to form true hyphae a variety of commonly used hypha inducing conditions were used. Assays were conducted in both static and shaking environments.

The first hyphal induction assay carried out consisted of induction with 10% FCS with a temperature switch from 30 °C (overnight growth) to 37 °C in 200 rpm shaking conditions. Samples were imaged at various time points and the results are shown in Figure 5.7. The WT AHY940 strain formed true hyphae readily in these conditions, while, as shown previously in Chapter 4, the *ΔΔtlo* mutant strain grew mainly as pseudohyphae. Reintroduction of *TLOα1* restored the ability to form true hyphae. This phenotype was apparently not affected by the level of expression of the gene as similar results were observed with both the *ENO1* and *TET1* promoters. While the *TLOβ2* reintroduction strains were already displaying true hyphal growth in some of the cells in the initial overnight culture, introduction into the hyphal inducing media was able to induce the switch to hyphae in the remaining yeast cells. Expression of *TLOγ11* in the *ΔΔtlo* background had no detectable effect on morphology and this strain grew as pseudohyphae, in a manner similar to the *ΔΔtlo* mutant strain. Similar data were also obtained using water supplemented with FCS as the hypha-inducing medium, Figure 5.8.

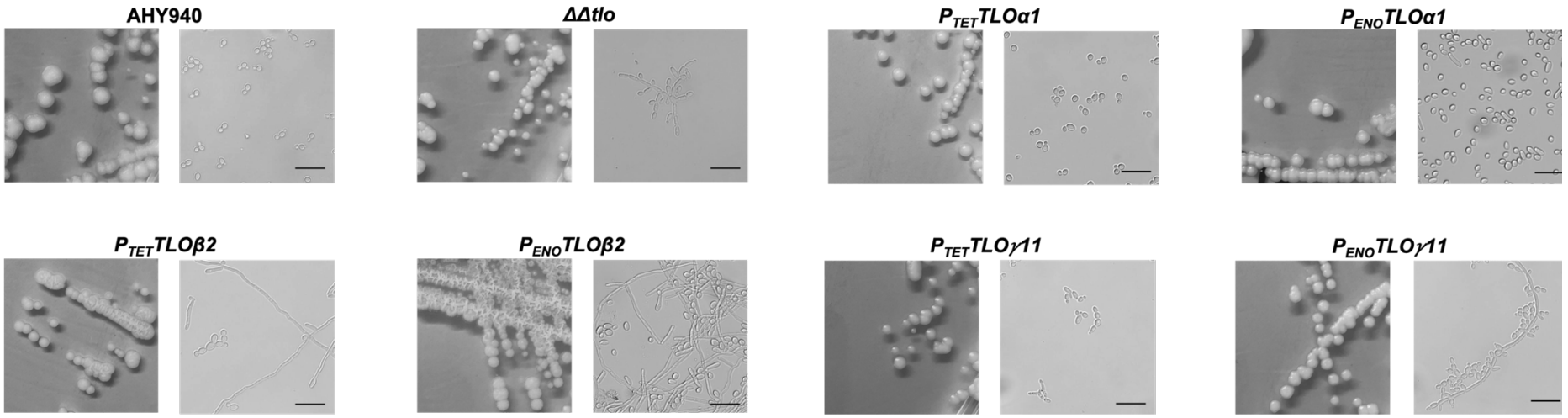


Figure 5.6 Colony and cellular morphology of *TLO* reintroduction strains

Colony morphology (left) was photographed after 48 h of growth on YEPD agar at 37 °C. Cellular morphology (right) was imaged under X400 magnification after 24 h of growth in YEPD liquid media. Scale-bar represents 7 μm.

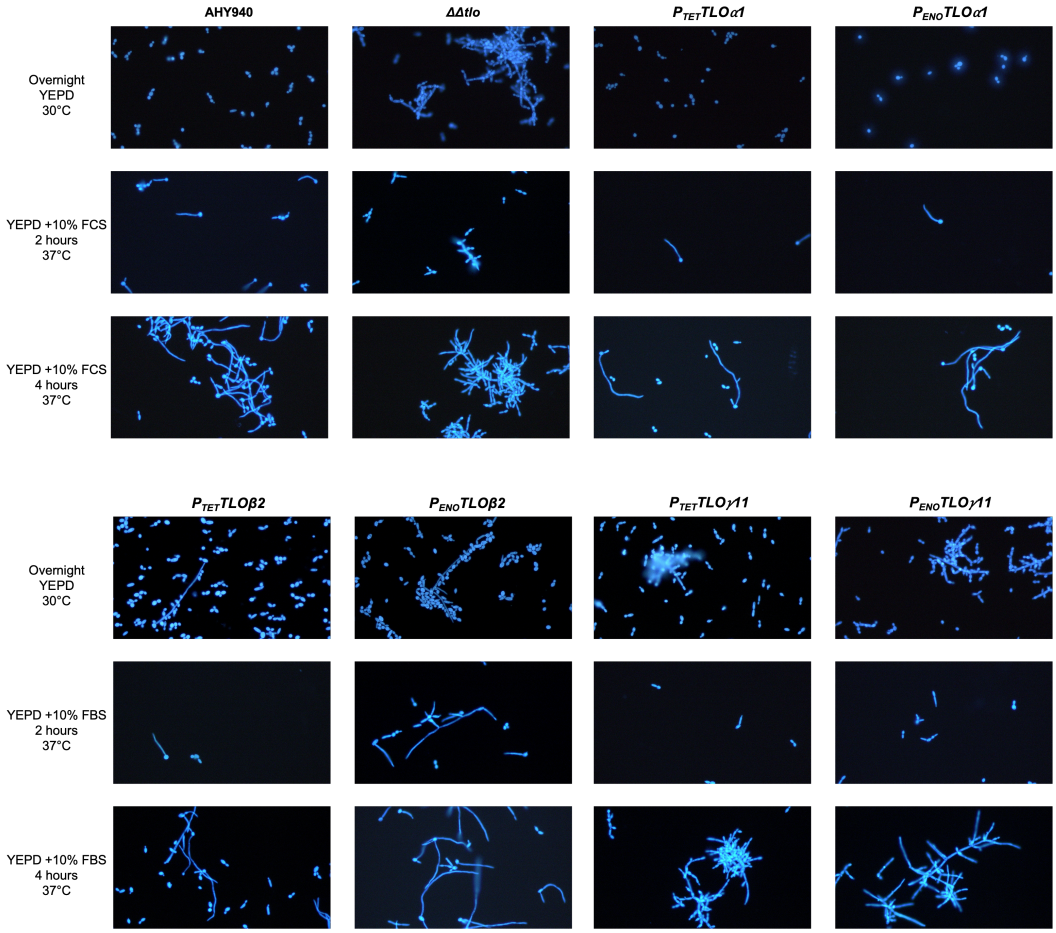


Figure 5.7 Hyphal Growth of *TLO* reintroduction strains in YEPD with 10% FCS

Flasks with 25 ml YEPD supplemented with 10% FCS were inoculated with 10^6 cfu/ml of each strain from cultures that had been grown overnight at 30 °C. Flasks were incubated at 37 °C in a 200 rpm shaking incubator. Cultures were monitored for hyphal formation. At each time point 10 µl of culture was removed and stained with Calcofluor White for imaging under the microscope.

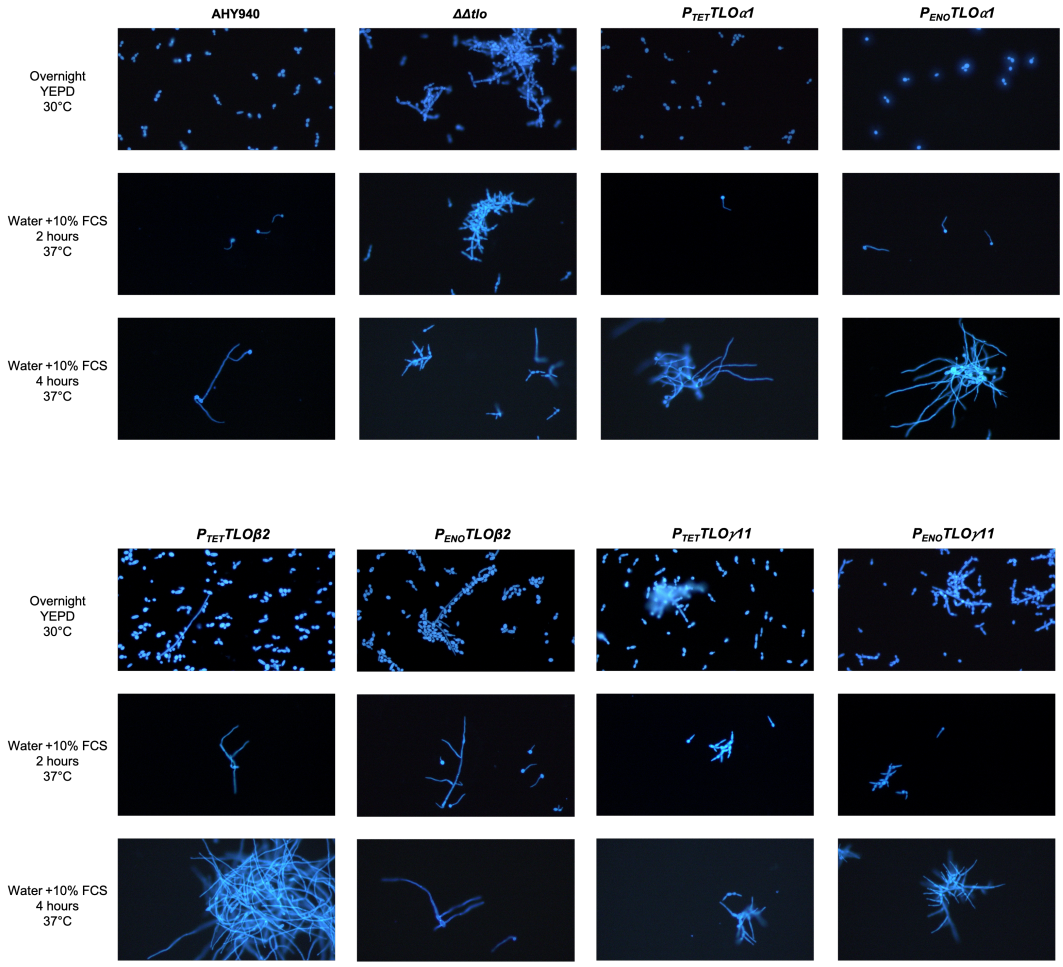


Figure 5.8 Hyphal Growth of *TLO* reintroduction strains in water with 10% FCS

Flasks with 25 ml water supplemented with 10% FCS were inoculated to 10^6 cfu/ml with each strain of from cultures that had been grown overnight at 30 °C. Flasks were incubated at 37 °C in a 200 rpm shaking incubator. Cultures were monitored for hyphal formation. At each time point 10 µl of culture was removed and stained with Calcofluor White for imaging under the microscope.

In the shaking assays, it was noted that a lot of cell clumping was occurring along the waterline in the culture flasks, so static assays were also performed and used to quantify the extent of hyphal formation in the strains. YEPD supplemented with 10% FCS was used as the induction medium, along with a temperature switch from an overnight culture grown at 30 °C to the assay being performed at 37 °C, Figure 5.9. The images in the figure were used to quantify the percentage of hyphal cells after overnight culture at 30 °C, and after 1 h in the hyphal inducing medium, Figure 5.10 (A, C and D), three separate fields of 100x100 μm were selected at random and the cell types within them counted. These figures were also used to estimate the length of hyphal elongation in the strains after 1 h in the inducing medium, Figure 5.10 (B and E). The results of this static assay were very similar to the results obtained when the samples were shaken, i.e. the WT AHY940 and *TLOα1* and *TLOβ2* reintroduction strains formed true hyphae under these conditions, while the *ΔΔtlo* mutant and the *TLOγ11* reintroduction strains grew primarily as pseudohyphae. There was a difference however in the percentages of hyphae formed. After overnight growth in YEPD at 30 °C with 200 rpm shaking, it was seen that the only strains which produced true hyphae to a significant extent were the *TLOβ2* reintroduction strains. Around 20% of the cells in the *ΔΔtlo::P_{ENO}TLOβ2* strains were true hyphae, and in the *ΔΔtlo::P_{TET}TLOβ2* strain about 6% of the cells grew as true hyphae. All other strains contained no true hyphae at this initial time point. After 1 h in the hyphal inducing media the AHY940 strain, *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains underwent the switch to hyphal growth, with 55-75% of cells growing as true hyphae. The *ΔΔtlo::P_{ENO}TLOα1* and *ΔΔtlo::P_{ENO}TLOβ2* strains, however, formed significantly more true hyphae than the WT AHY940 strain, with 80% of cells growing as hyphae in the *ΔΔtlo::P_{ENO}TLOα1* strain and around 90% of the cells in *ΔΔtlo::P_{ENO}TLOβ2* growing as hyphae. There was a significant difference in the length of the hyphae formed at 1 h in this experiment by the *ΔΔtlo::P_{ENO}TLOβ2* strain compared with the lengths of the hyphae in all the other strains. In the AHY940, *ΔΔtlo::P_{TET}TLOα1, ΔΔtlo::P_{ENO}TLOα1* and the *ΔΔtlo::P_{TET}TLOβ2* strains at 1 h post inoculation into the inducing media the hyphae were estimated to measure between 9.5-11.5 μm, but the hyphae formed by the *ΔΔtlo::P_{ENO}TLOβ2* strain were estimated to be around 15 μm, and determined to be significantly longer than any of the other strains. In this analysis the *ΔΔtlo* and the *TLOγ11* reintroduction strains did not grow as true hyphae. Although their pseudohyphal protrusions elongated, their lengths were not compared to the truly hyphal strains.

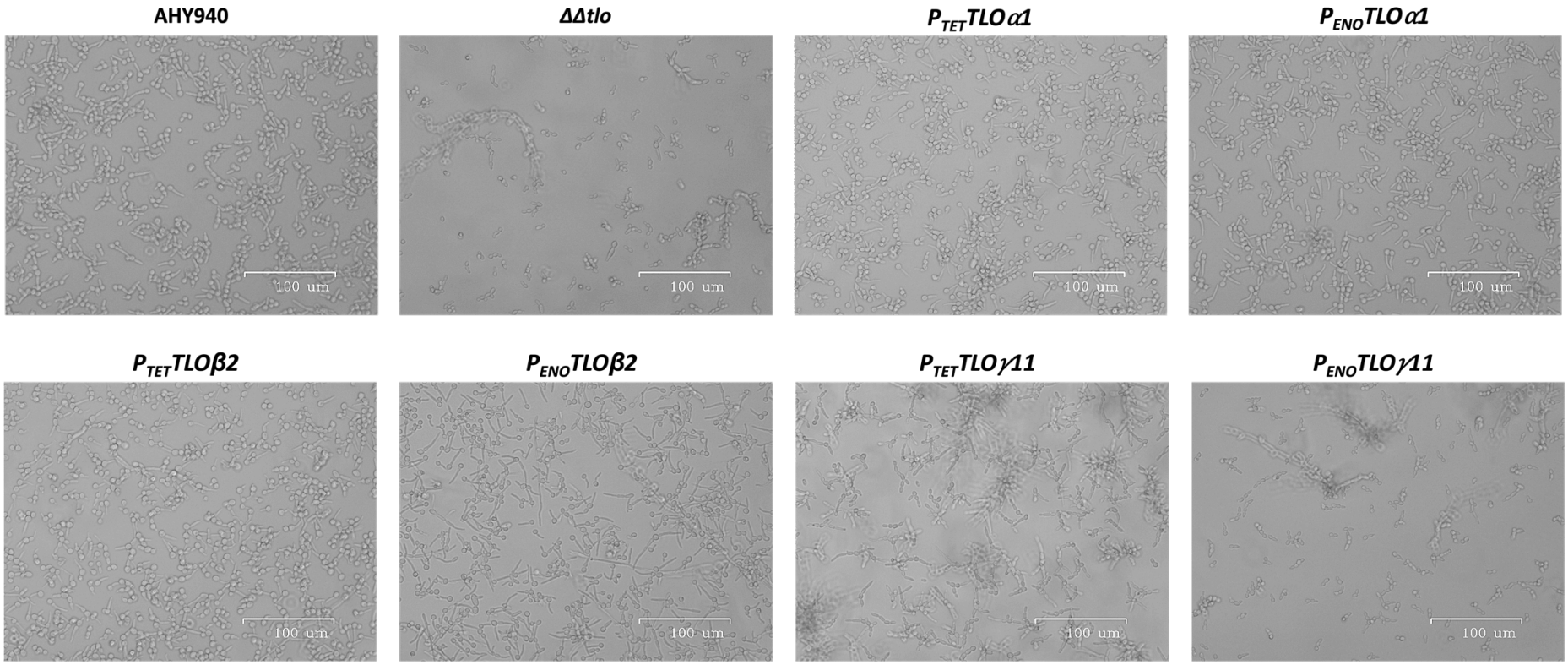
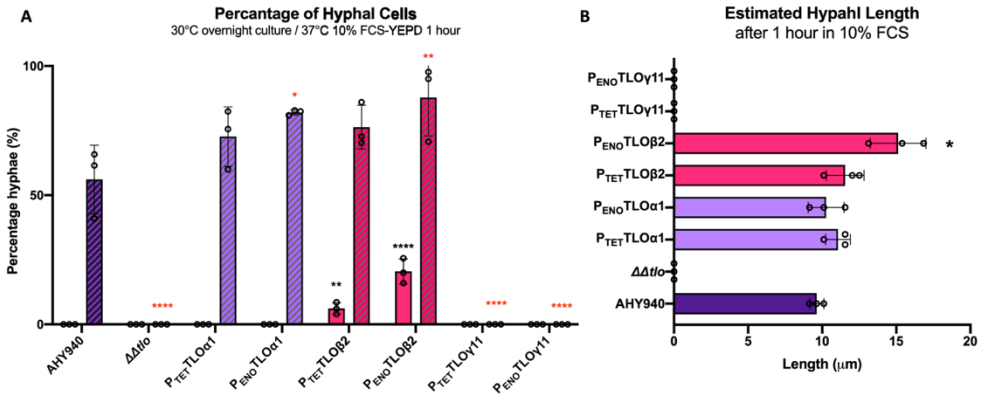


Figure 5.9 Hyphal Growth of *TLO* reintroduction strains in YEPD with 10% FCS in static growth conditions

Each well of a 6-well plate was set up with 2 ml YEPD with 10% FCS were inoculated to an OD_{600} of 0.2 with the required strain from an overnight culture grown at 30 °C with 200 rpm shaking. Flasks were incubated at 37 °C in a static incubator. Plates were removed from the incubator at 1 and imaged with the Zoe Fluorescent Cell Imager without removing lids.



C

	AHY940	*Δ/Δtlo*	P_{TET}TLOα1	P_{ENO}TLOα1	P_{TET}TLOβ2	P_{ENO}TLOβ2	P_{TET}TLOγ11	P_{ENO}TLOγ11
Mean (%) at 0 hours	0.000	0.000	0.000	0.000	6.167	20.53	0.000	0.000
S.D.	0.000	0.000	0.000	0.000	2.255	4.822	0.000	0.000
Mean (%) at 1 hour	56.10	0.000	72.67	82.03	76.33	87.80	0.000	0.000
S.D.	13.25	0.000	11.48	1.002	8.457	14.87	0.000	0.000
Mean (μm) at 1 hour	9.626	0.000	11.07	10.27	11.55	15.12	0.000	0.000
S.D.	0.4830	0.000	0.8314	1.211	1.270	1.873	0.000	0.000

Figure 5.10 Quantification of hyphal growth of *TLO* reintroduction strains in YEPD with 10% FCS in static growth conditions

The percentage of true hyphal growth was determined for each strain by capturing three fields of view and counting all cells; noting if cells were either hyphal, pseudohyphal or yeast. The percentage of cells that were hyphal was calculated and graphed. (A) Graphical representation of percentage of cells that are truly hyphal in either an overnight culture in YEPD grown at 30 °C with 200 rpm shaking (solid bars) or in static YEPD with 10% FCS at 37 °C (hashed bars) (B) Estimation of the length of true hyphae after 1 hour growth in static YEPD with 10% FCS at 37 °C. Lengths were measured relative to the scale bar 100 μm² field, these measurements were performed for 8-10 hyphae in each field and performed in three separate fields. (C) Mean and standard deviation of results in A and B.

To determine the ability of the strains to form hyphae on solid media, the strains were grown as streaks on Spider agar at either 30 or 37 °C for five days, Figure 5.11. It was seen that at 30 °C the AHY940 strain presented with a wrinkled form with a hyphal fringe evident. The *ΔΔtlo* strain did not display any wrinkling and showed minimal fringing. The *TLOα1* reintroduction strains both grew as smooth streaks, but did show some formation of a hyphal fringe. The *TLOβ2* reintroduction strains grew as highly wrinkled streaks on the plates, more so than the WT, and also displayed a hyphal fringe. With the *TLOα1* and *TLOβ2* reintroduction strains there did not appear to be any difference between the strains expressing the genes at different levels from the *ENO1* or *TET1* promoters. For the *TLOγ11* reintroduction strains there was a difference in the effect of high and low expression on phenotype. The *ΔΔtlo::*P_{TET}*TLOγ11* strain grew similarly to the *ΔΔtlo* mutant, with a smooth morphology and minimal hyphal fringe. The *ΔΔtlo::*P_{ENO}*TLOγ11* strain however, displayed a more wrinkled phenotype and hyphal fringe. At 37 °C, similar patterns of morphology were observed. There appeared to be a greater degree of wrinkling in the WT and *TLOβ2* reintroduction strains at 37 °C compared to 30 °C. The *TLOα1* reintroduction strains grew similarly at 37 °C as they did at 30 °C. The *ΔΔtlo* mutant and the *ΔΔtlo::*P_{TET}*TLOγ11* strains appeared more wrinkled at 37 °C than 30 °C, with less apparent hyphal fringing. The growth of *ΔΔtlo::*P_{ENO}*TLOγ11* at 37 °C was very different to the growth at 30 °C, appearing less wrinkled, with less hyphal fringing, resembling the *ΔΔtlo* mutant and the *ΔΔtlo::*P_{TET}*TLOγ11* strain. There did appear to be more hyphal fringe forming in the *ΔΔtlo::*P_{ENO}*TLOγ11* strain than the *ΔΔtlo* mutant and the *ΔΔtlo::*P_{TET}*TLOγ11* strains. There did not appear to be a difference in phenotype between the strains expressing the same gene at different levels when grown at 37 °C. Spider plate assays were also set up where a spot of each strain (5 µl of 10^6 cfu/ml) was grown on Spider agar for 5 days at 37 °C, Figure 5.12. This was performed to better examine the hyphal fringing morphology of the strains. The strains here followed a similar pattern to the streaked Spider plates. Differences became apparent between the AHY940 WT strain and the *TLOβ2* reintroduction strains, with the inner colony appearing more wrinkled in the *TLOβ2* reintroduction strains, but hyphal fringing being less intense, though the *ΔΔtlo::*P_{TET}*TLOβ2* strain resembled the WT more than *ΔΔtlo::*P_{ENO}*TLOβ2.* Differences between the *TLOγ11* reintroduction strains also became more evident, with the *ΔΔtlo::*P_{TET}*TLOγ11* strain behaving like the *ΔΔtlo* mutant, while the *ΔΔtlo::*P_{ENO}*TLOγ11,* although displaying a similar central region, displayed a much more pronounced hyphal fringe.

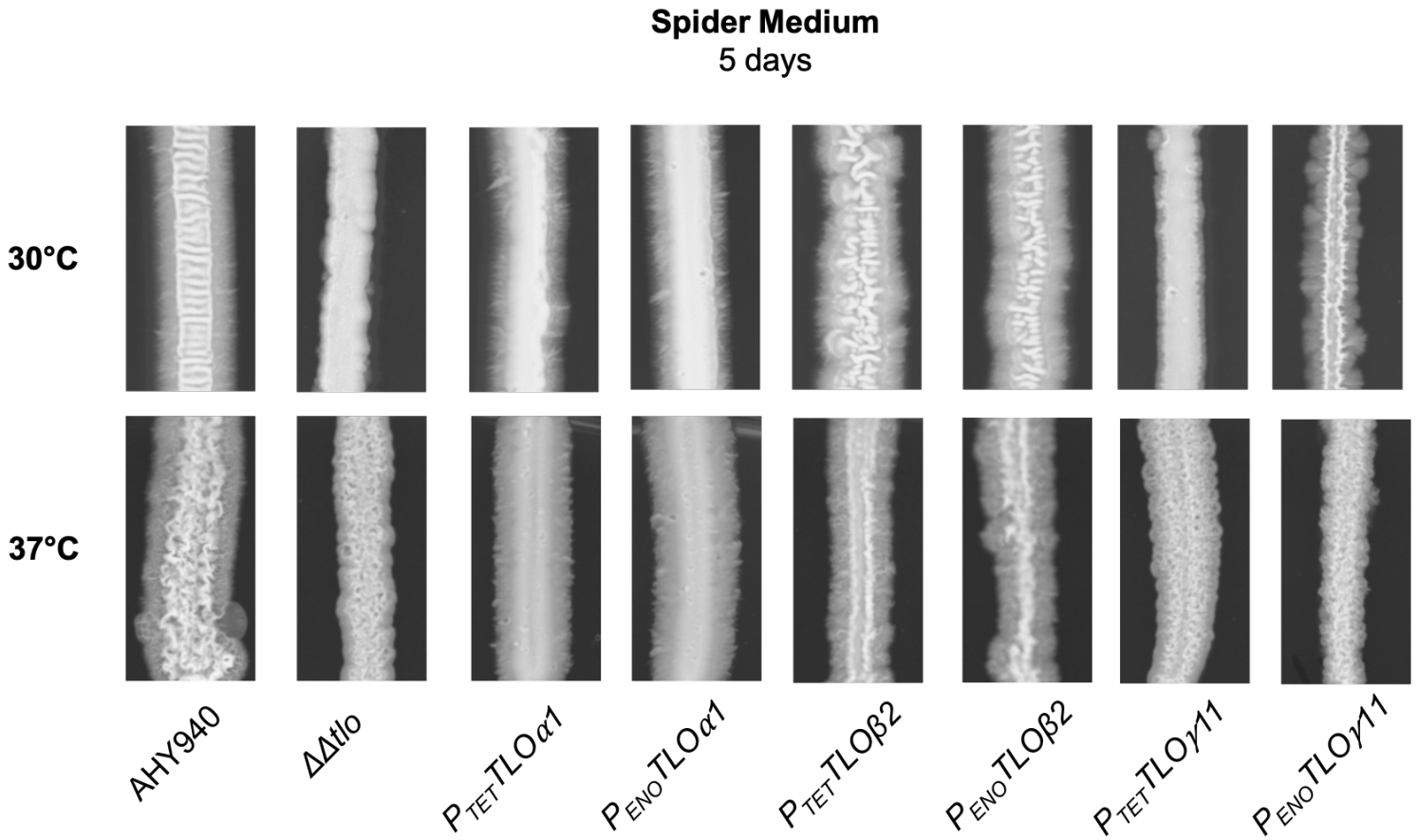


Figure 5.11 Growth of *TLO* reintroduction strains on solid Spider media

The ability of strains to filament on solid hyphal media was determined by streaking a single colony of each strain onto an agar plate and incubating at either 30 °C or 37 °C for 5 days. After 5 days plates were examined and photographed.

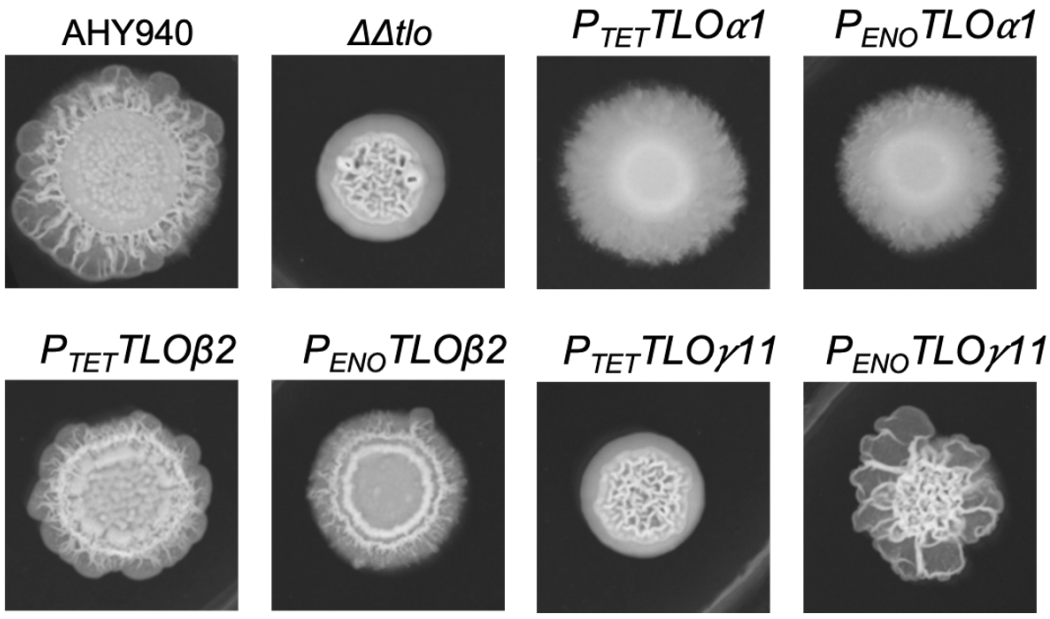


Figure 5.12 Morphology of *TLO* reintroduction strains on solid Spider media

An overnight culture of each strain was standardised to a concentration of 10^6 cfu/ml. 5 µl of each suspension was spotted onto Spider agar and allowed to dry, a maximum of three spots were placed on each plate to avoid crowding. These plates were incubated for 5 days at 37 °C. After 5 days plates were examined and photographed.

5.3.5.3 Chlamydospore formation

The ability of the strains to form chlamydospores on inducing media was tested, Figure 5.13. As described previously, the WT AHY940 strain formed chlamydospores, and the *ΔΔtlo* strain did not. Reintroduction of *TLOα1* or *TLOβ2*, but not *TLOγ11,* restored the ability of the cells to form chlamydospores. The level of expression of the *TLO* genes had no effect on this phenotype.

5.3.5.3 Growth rate analysis

To determine if reintroduction of *TLO* genes into the *ΔΔtlo* background had an effect on growth, the growth rates of the strains in both YEPD and YEP-Galactose at 37 °C were determined, Figure 5.13.

In YEPD, it was seen that the *ΔΔtlo* strain grew much more slowly than the parental WT AHY940 strain. Reintroduction of *TLOα1* or *TLOβ2*, whether expressed from the *ENO1* or the *TET1* promoter, restored the growth to a rate similar to that of the WT. Introduction of *TLOγ11* from the *TET1* promoter did not have any effect on the lower growth rate of the *ΔΔtlo* mutant strain, but expression of *TLOγ11* from the *ENO1* promoter was able to restore the growth rate to that of the WT. Similar results were obtained when YEP-Galactose was used as the culture medium, with the exception that introduction of *TLOγ11* from the *ENO1* promoter had no effect on growth rate, and introduction of *TLOγ11* from the *TET1* promoter resulted in a decrease in growth rate compared to the *ΔΔtlo* mutant.

It was noted that cellular morphology was affected when the cells were grown in YEP-Galactose, specifically that the *TLOβ2* reintroduction strains grew exclusively in the yeast morphology and no hyphae were detectable when expressed from both the *ENO1* and *TET1* promoters, Figure 5.15. The morphology of all other strains was the same in YEPD as in YEP-Galactose.

The growth of strains on YEPD agar at a range of temperatures (22-42 °C) was also examined using a spot plate assay, Figure 5.16. All strains and mutants tested grew to a similar degree, however, the wrinkled morphology of the *TLOβ2* reintroduction strains at 30, 37 and 42 °C disappeared when grown at 22 °C, and the colonies more resembled those of the other strains.

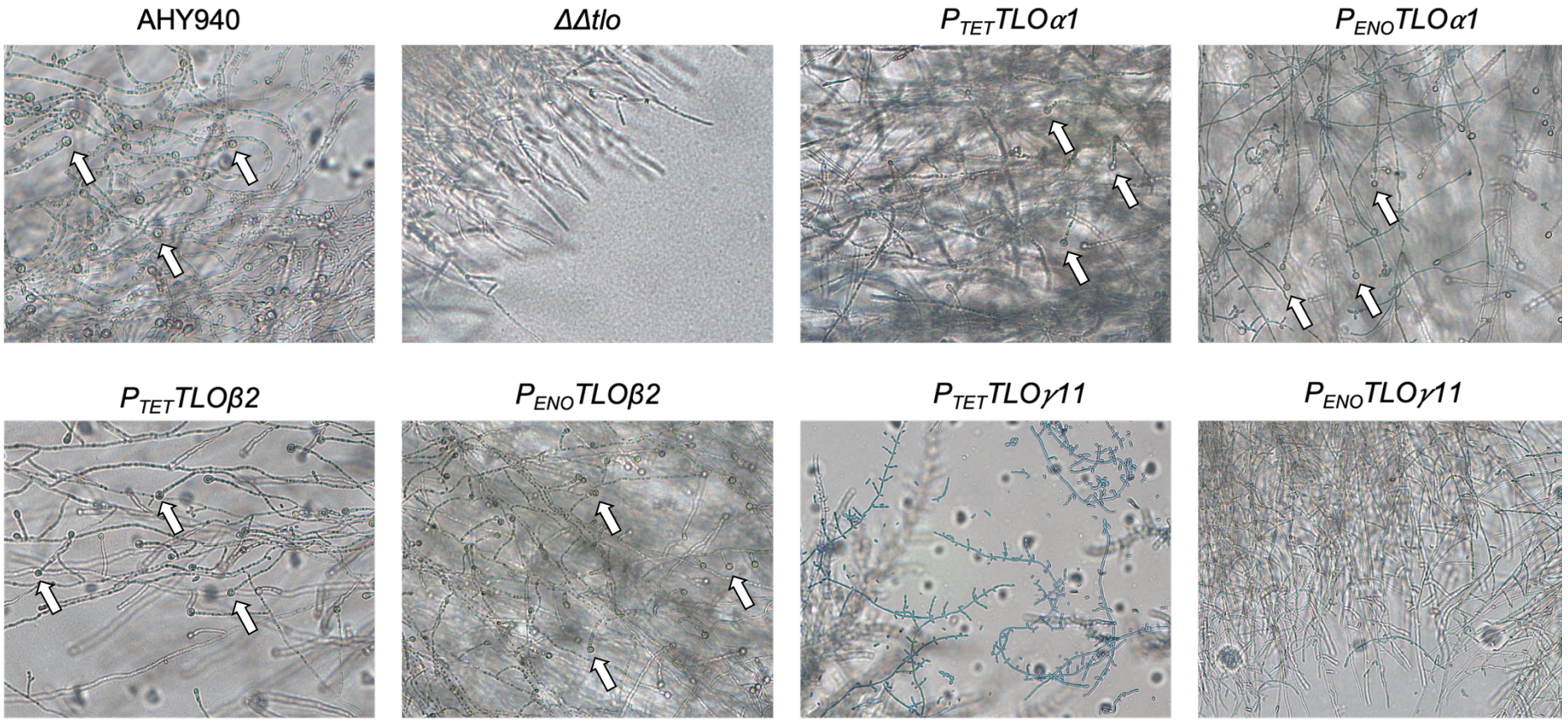
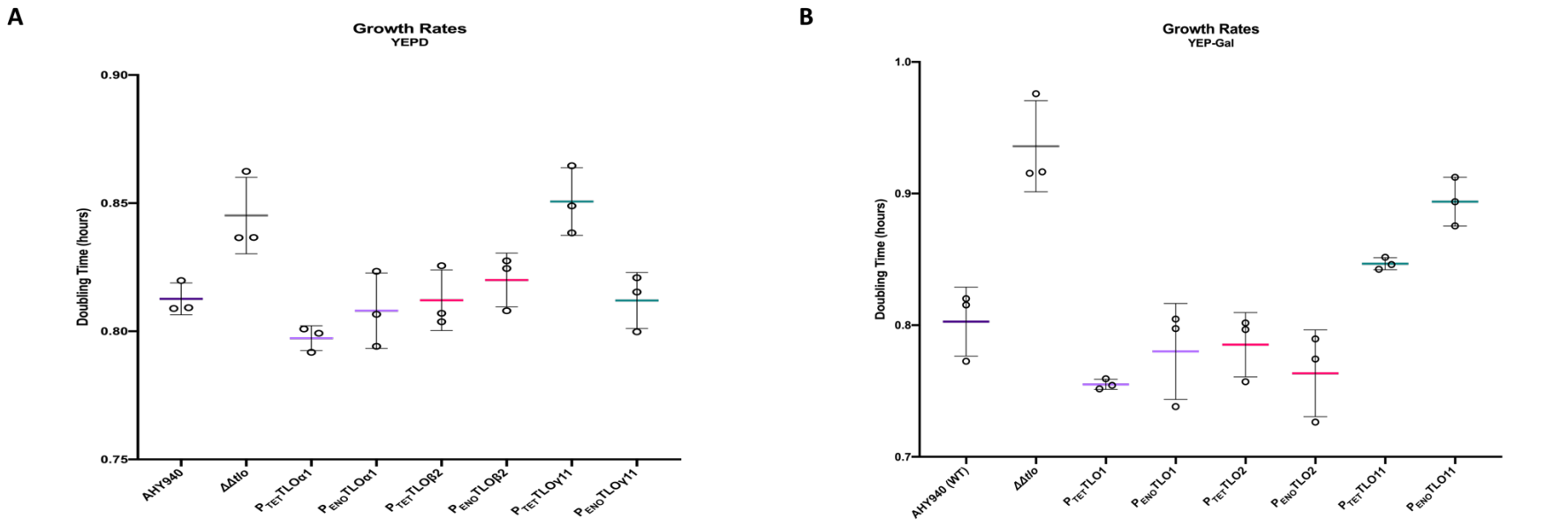


Figure 5.13 Chlamydospore Formation of *TLO* reintroduction strains on Corn Meal Agar-Tween80

A single colony of each strain was streaked lightly onto corn meal agar supplemented with 1% Tween 80. Streaks were covered with a glass coverslip and the plates were incubated for 5-7 days at 22 °C in the dark. After incubation the plates were viewed and imaged directly under a light microscope. Exemplary chlamydospores are indicated by white arrows.



C

YEPD	AHY940	*ΔΔtlo*	P_{TET}TLOα1	P_{ENO}TLOα1	P_{TET}TLOβ2	P_{ENO}TLOβ2	P_{TET}TLOγ11	P_{ENO}TLOγ11
Mean (h)	0.8126	0.8452	0.7973	0.808	0.8121	0.82	0.8506	0.812
S.D.	0.006208	0.01492	0.004838	0.0147	0.01181	0.0105	0.01319	0.01093
Significantly different to *ΔΔtlo*	*	-	***	**	*	ns	ns	*
P-value	0.0164	-	0.0006	0.006	0.0146	0.0769	0.9874	0.0143
Significantly different to AHY940	-	*	ns	ns	ns	ns	**	ns
P-value	-	0.0164	0.4407	0.9945	>0.9999	0.943	0.005	>0.9999

D

YEP-Gal	AHY940	*ΔΔtlo*	P_{TET}TLOα1	P_{ENO}TLOα1	P_{TET}TLOβ2	P_{ENO}TLOβ2	P_{TET}TLOγ11	P_{ENO}TLOγ11
Mean (h)	0.8027	0.9359	0.7551	0.7801	0.7852	0.7634	0.8467	0.8939
S.D.	0.02621	0.03462	0.003907	0.03647	0.02443	0.03297	0.004583	0.0185
Significantly different to *ΔΔtlo*	****	-	****	****	****	****	**	ns
P-value	<0.0001	-	<0.0001	<0.0001	<0.0001	<0.0001	0.0035	0.2583
Significantly different to AHY940	-	****	ns	ns	ns	ns	ns	**
P-value	-	<0.0001	0.1658	0.8061	0.9262	0.3171	0.2226	0.0029

Figure 5.14 Growth rate analysis of *TLO* reintroduction strains

Doubling time of each strain is represented by horizontal line (hours), with error bars representing standard deviation and symbols representing each of three replicates. (A) Doubling times of strains growing in liquid YEPD at 37 °C at 200 rpm. (B) Doubling times of strains growing in liquid YEP-Galactose at 37 °C at 200 rpm. (C) Statistical analysis of data from strains growing in YEPD. A one-way ANOVA was performed to determine if results were significantly different and a Dunnett's multiple comparisons test was performed to determine which means were significantly different from WT AHY940, and another to determine significant difference from the ΔΔtlo *s*train, asterisks represent degrees of significance. (D) As with C, represents statistical analysis of strains growing in YEP-Galactose.

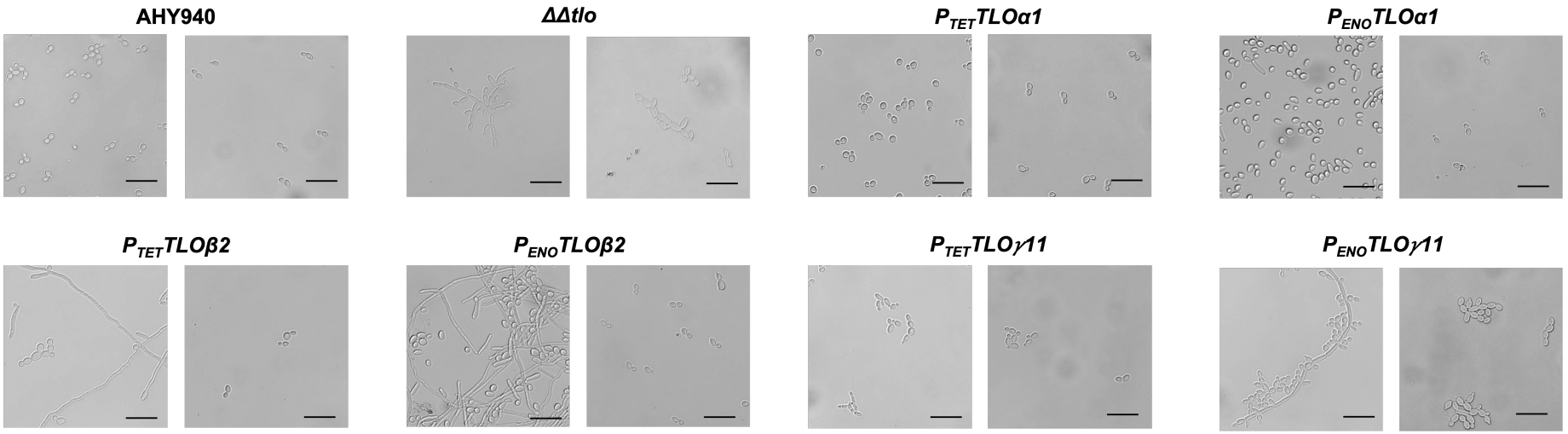


Figure 5.15 Morphology of *TLO* reintroduction strains in different carbon sources

The morphology of cells growing in YEPD (left) and in YEP-Galactose (right) were imaged using a light microscope. Cells were taken from an overnight culture of YEPD grown at 37 °C with 200 rpm shaking and imaged. These overnight cultures were used to inoculate a fresh YEP-Galactose broth and grown for 6 hours, after which time, a sample of the cells were removed and imaged. Scale-bar represents 7 μm.

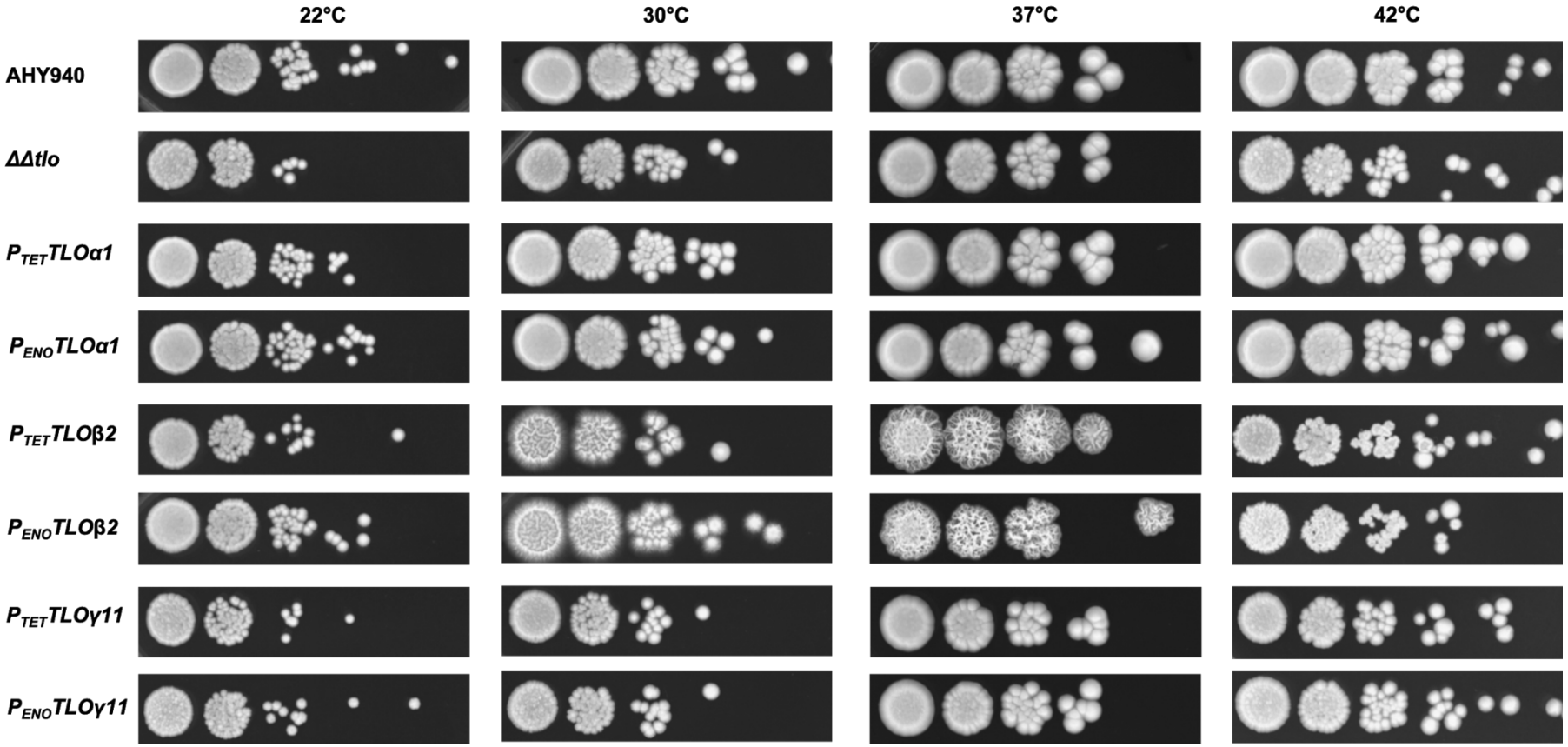


Figure 5.16 Growth of *TLO* reintroduction strains on YEPD at various temperatures

10-fold serial dilutions (10^6 – 10^1 cfu/ml) were plated in 5 µl volumes onto YEPD agar and incubated for 72 h at 22, 30, 37 or 42 °C before photographing.

The ability of the strains to grow on YEP-Galactose in the presence of a respiration inhibitor, Antimycin A, was examined by spot plate assay to determine if the strains could metabolise carbon from galactose when respiration was inhibited, Figure 5.17. Growth on YEP-Galactose only was used as a control, but also highlighted differences between strains. The *TLOα1* reintroduction strains displayed a wrinkled colony morphology, but more closely resemble the morphology of the WT than any other strain. The *TLOγ11* reintroduction strains closely resembled the *ΔΔtlo* mutant, with reduced growth and a smooth colony morphology. The *TLOβ2* reintroduction strains were more intense in their wrinkled morphology, again with slight difference between the high and low expressing strains. When these strains were grown in liquid YEP-Galactose medium previously, the morphologies did not appear hyphal, Figure 5.15, but two main differences in cellular morphology were seen, i.e. yeast and pseudohyphal morphologies. The strains that grew in the yeast morphology in the broth were the same as those that grew wrinkled on the agar, AHY940, the *TLOα1* and the *TLOβ2* reintroduction strains. Those that grew pseudohyphally in the broth were those that grew as smooth colonies on the agar, the *ΔΔtlo* and the *TLOγ11* reintroduction strains.

In the presence of Antimycin A, the *TLOα1* and *TLOβ2* reintroduction strains all grew similarly to the WT. The *ΔΔtlo::*P_{TET}*TLOγ11* strain behaved like the *ΔΔtlo* strain, however, the *ΔΔtlo::*P_{ENO}*TLOγ11* strain showed a restoration of growth and resembled the WT strain.

5.3.5.3 Biofilm formation

To determine if reintroduction of *TLOs* into the *ΔΔtlo* background had an effect on biofilm formation, the ability of the strains to form biofilm on plastic surfaces was measured, Figure 5.18. At the 24 h time point, all strains formed more biofilm than the *ΔΔtlo* strain with the exception of the *ΔΔtlo::*P_{TET}*TLOγ11* strain, which grew similarly to the *ΔΔtlo* strain. The *ΔΔtlo::TLOα1* strain and both of the *TLOβ2* reintroduction were significantly different from the *ΔΔtlo* mutant. At this time point, the *ΔΔtlo::*P_{ENO}*TLOγ11* strain formed a substantially higher level of biofilm than all the other strains, including the AHY940 WT. At the 48 h time point, the level of biofilm formed by the *ΔΔtlo::*P_{ENO}*TLOγ11* strain was similar to that of the WT and the other strains tested, with the exception of the *ΔΔtlo::*P_{TET}*TLOγ11* reintroduction strain, which produced similar levels of biofilm as the *ΔΔtlo* mutant strain. The *TLOα1* reintroduction strains also formed significantly more biofilm than the AHY940 WT.

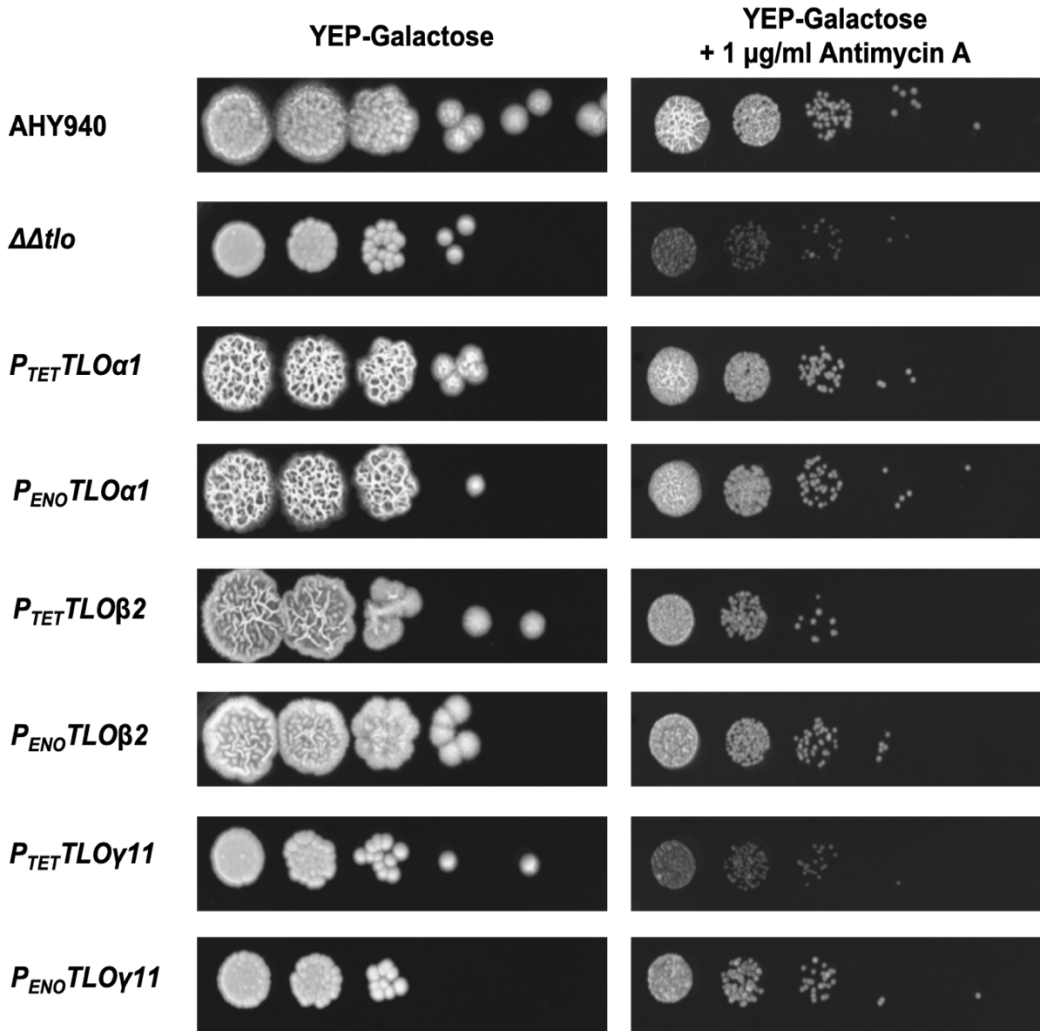
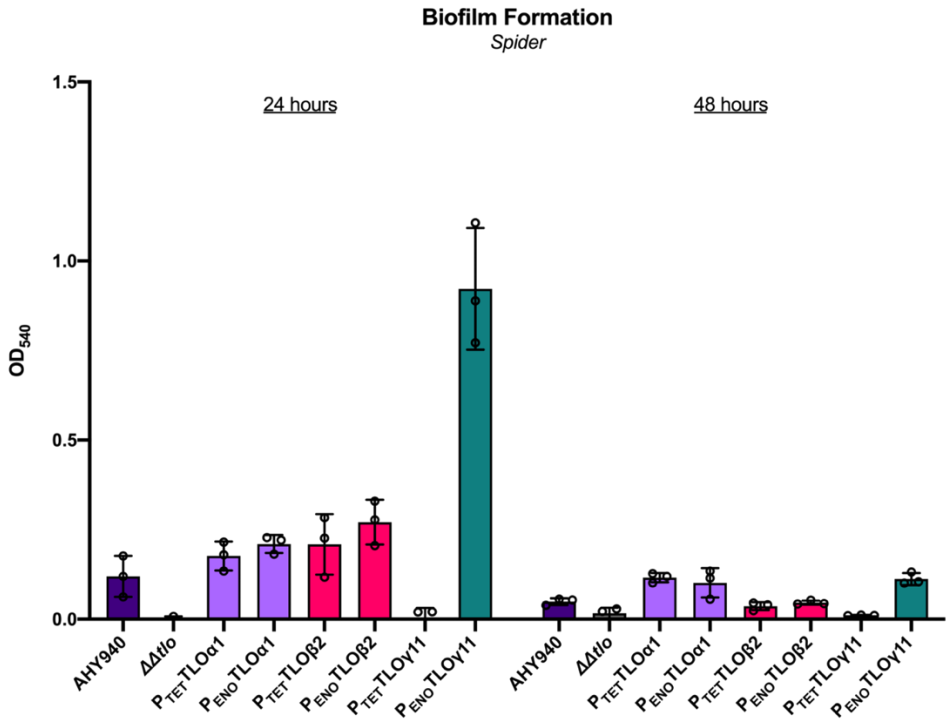


Figure 5.17 Growth of *TLO* reintroduction strains in the presence of Antimycin A

10-fold serial dilutions (10^6 – 10^1 cfu/ml) were plated in 5 µl volumes onto YEP-Galactose agar with and without 1 µg/ml Antimycin A and incubated for 72h at 37 °C before photographing.

A



B

24 h	AHY940	ΔΔtlo	P_{TET}TLOα1	P_{ENO}TLOα1	P_{TET}TLOβ2	P_{ENO}TLOβ2	P_{TET}TLOγ11	P_{ENO}TLOγ11
Mean	0.1195	-0.004767	0.1766	0.21	0.209	0.271	0.004833	0.9223
S.D.	0.0573	0.0128	0.04063	0.0249	0.08434	0.06229	0.02702	0.1699
Significantly different to ΔΔtlo	ns	-	ns	*	*	**	ns	****
P-value	0.2605	-	0.0505	0.0174	0.018	0.0024	0.9998	<0.0001
Significantly different to AHY940	-	ns	ns	ns	ns	ns	ns	****
P-value	-	0.2605	0.8924	0.5585	0.5688	0.1241	0.3306	<0.0001

C

48 h	AHY940	ΔΔtlo	P_{TET}TLOα1	P_{ENO}TLOα1	P_{TET}TLOβ2	P_{ENO}TLOβ2	P_{TET}TLOγ11	P_{ENO}TLOγ11
Mean	0.04931	0.01652	0.1161	0.1016	0.03661	0.04656	0.01101	0.1123
S.D.	0.009323	0.01588	0.0131	0.04096	0.01117	0.005671	0.0007493	0.0168
Significantly different to ΔΔtlo	ns	-	****	***	ns	ns	ns	****
P-value	0.1829	-	<0.0001	0.0002	0.6247	0.2488	0.9994	<0.0001
Significantly different to AHY940	-	ns	**	*	ns	ns	ns	**
P-value	-	0.1829	0.0021	0.0149	0.9183	0.9997	0.0944	0.0034

Figure 5.18 Biofilm formation of *TLO* reintroduction strains

The ability of strains to form biofilm on plastic surfaces (plastic 96 well plate) in liquid Spider media was analysed. Biofilm formation was quantified by staining with crystal violet and measuring absorbance at 540 nm. This was done at 24 and 48 hour time points. (A) Graphical representation of biofilm formation data. Error bars represent standard deviation from mean, and symbols represent values from three replicates. (B&C) Statistical analysis by means on one way ANOVA of 24 h and 48 h time point data respectively.

5.3.5.4 Cell wall perturbing compound spot plate assays

Spot plate assays were carried out to determine if the susceptibility of strains to cell wall perturbing compounds Calcofluor White and Congo Red was altered by introduction of *TLO*s into the null mutant background, Figure 5.19. The *ΔΔtlo* mutant strain was unable to grow in the presence of 25 μg/ml Congo Red. However, the *TLOα1* and *TLOβ2* reintroduction strains restored growth and grew at levels similar to the WT. The *ΔΔtlo::*P_{TET}*TLOγ11* strain grew similarly to the *ΔΔtlo* mutant strain, however, the *ΔΔtlo::*P_{ENO}*TLOγ11* strain showed increased growth, but not quite to the same degree as the WT. When the concentration of Congo Red was increased to 50 μg/ml the *ΔΔtlo::*P_{ENO}*TLOγ11* reintroduction strain failed to grow.

When cultured on agar containing 150 μg/ml Calcofluor White, all mutant strains showed reduced growth when compared to WT. The *TLOα1* and *TLOβ2* reintroduction strains were slightly more sensitive than WT. No growth was observed with the *ΔΔtlo::*P_{TET}*TLOγ11* strains, however, the *ΔΔtlo::*P_{ENO}*TLOγ11* reintroduction strain exhibited a limited degree of growth. When the concentration was increased to 300 μg/ml, only the WT and the *ΔΔtlo::*P_{ENO}*TLOα1* and *ΔΔtlo::*P_{ENO}*TLOβ2* strains displayed any growth. There was no growth of the *ΔΔtlo* strains, the *ΔΔtlo::*P_{TET}*TLOα1, ΔΔtlo::*P_{TET}*TLOβ2* or of either of the *TLOγ11* reintroduction strains.

5.3.5.5 Oxidative stress minimum inhibitory concentration assays

Microbroth dilution assays were performed in order to examine the sensitivity of these strains to oxidative stress inducing compounds, H_2O_2 and tBOOH (*tert*-Butyl hydroperoxide), Figure 5.20. Resistance to H_2O_2 induced oxidative stress was restored to the WT level in the *TLOα1* and *TLOβ2* reintroduction strains, with the highly expressing *TLOβ2* strain being more resistant than WT. Expression of the *TLOγ11* gene from either the *TET1* or *ENO1* promoter had no effect on oxidative stress tolerance.

Tolerance of oxidative stress induced with tBOOH was also examined. As with H_2O_2, the *ΔΔtlo* strain is much more sensitive than the WT strain. Reintroduction of *TLOα1* into the *ΔΔtlo* background under either promoter restored tolerance to the same level as the WT. Reintroduction of *TLOγ11* did not increase tolerance of the null background to tBOOH while reintroduction of *TLOβ2* under either promoter resulted in IC_{50} measurements intermediate between those of the WT and the *ΔΔtlo*.

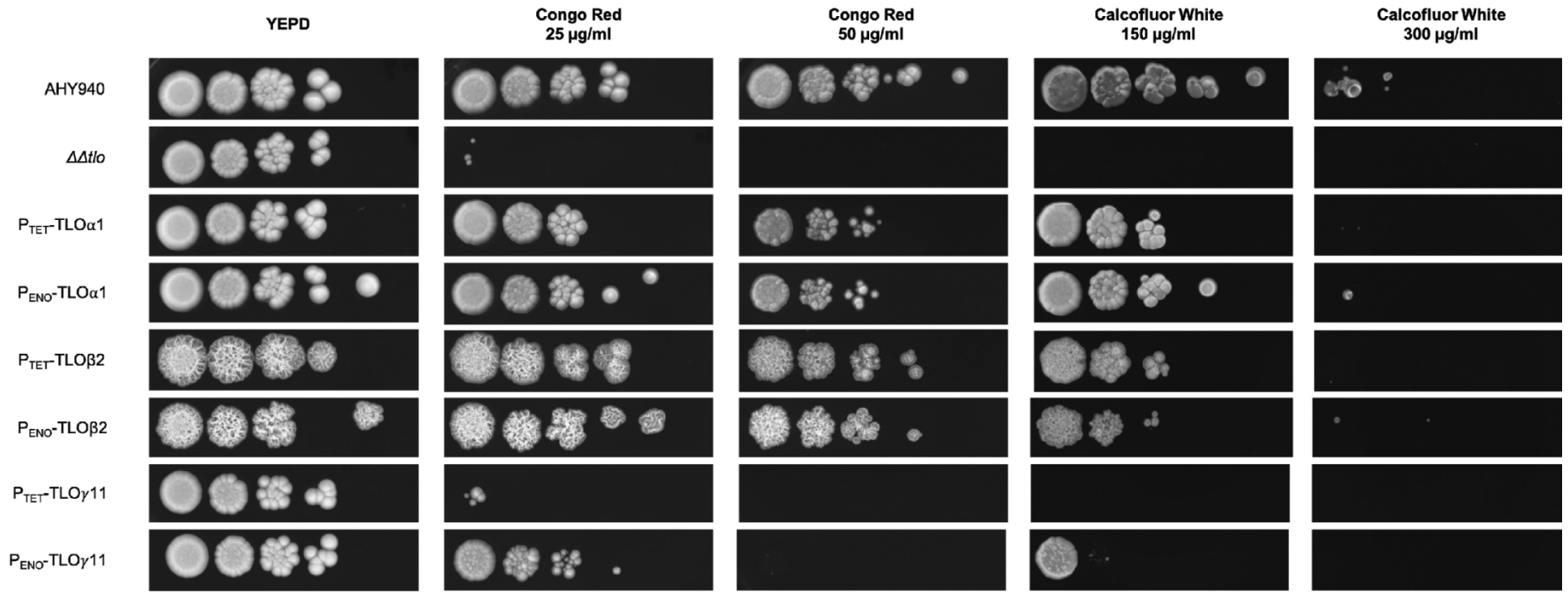
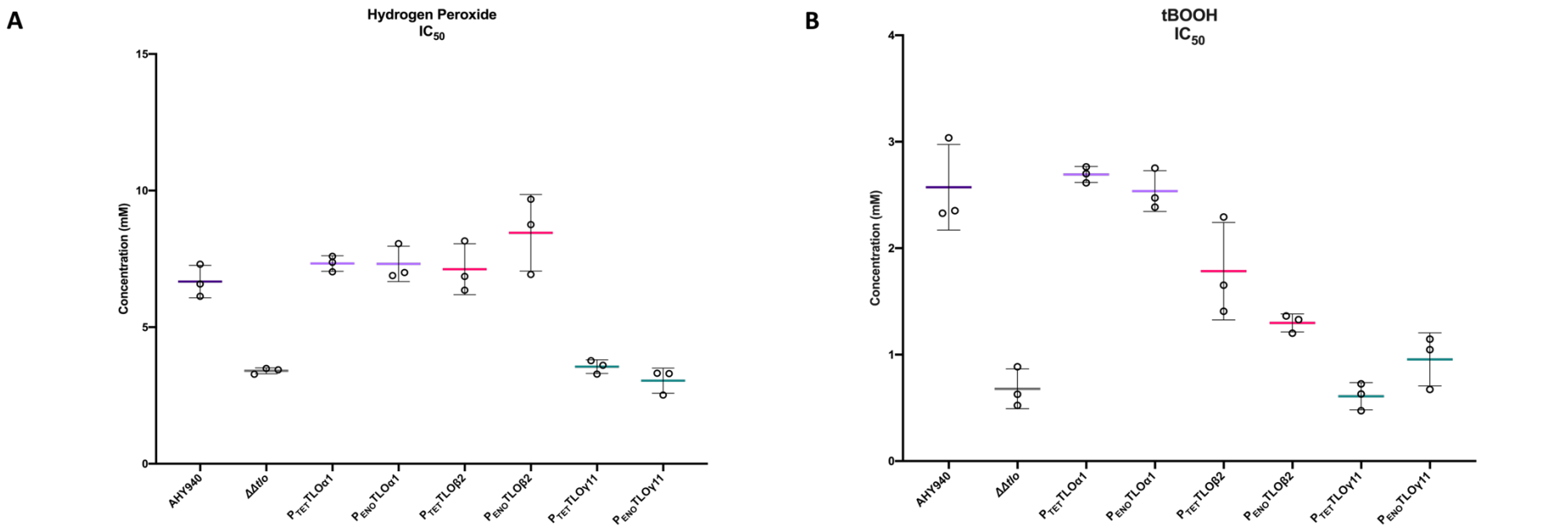


Figure 5.19 Growth of *TLO* reintroduction strains on media containing cell wall perturbing compounds

10-fold serial dilutions (10^6 – 10^1 cfu/ml) were plated in 5 µl volumes onto YEPD agar containing different concentrations of either Congo Red or Calcofluor White to determine susceptibility of strains to these compounds. Plates were incubated at 37 °C for 72h before photographing.



C

H_2O_2	AHY940	*ΔΔtlo*	P_{TET}TLOα1	P_{ENO}TLOα1	P_{TET}TLOβ2	P_{ENO}TLOβ2	P_{TET}TLOγ11	P_{ENO}TLOγ11
Mean (mM)	6.671	3.401	7.331	7.317	7.121	8.454	3.555	3.043
S.D.	0.5929	0.1087	0.2862	0.6456	0.9318	1.401	0.249	0.4587
Significantly different to ΔΔtlo	***	-	****	****	****	****	ns	ns
P-value	0.0002	-	<0.0001	<0.0001	<0.0001	<0.0001	0.9996	0.9819
Significantly different to AHY940	-	***	ns	ns	ns	*	***	****
P-value	-	0.0002	0.7635	0.7789	0.9452	0.0353	0.0004	<0.0001

D

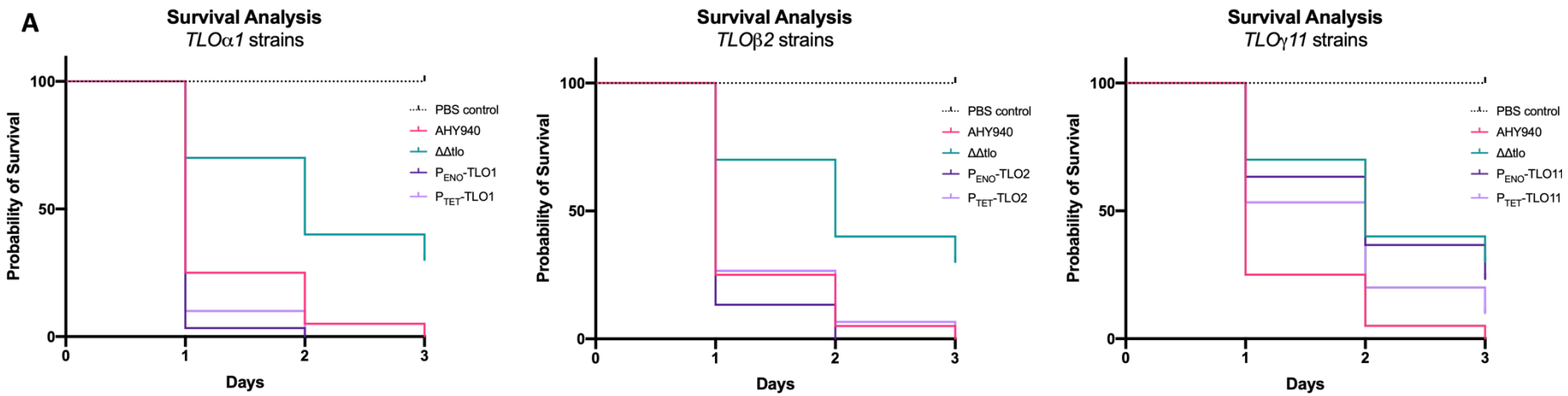
tBOOH	AHY940	*ΔΔtlo*	P_{TET}TLOα1	P_{ENO}TLOα1	P_{TET}TLOβ2	P_{ENO}TLOβ2	P_{TET}TLOγ11	P_{ENO}TLOγ11
Mean (mM)	2.573	0.6793	2.693	2.537	1.784	1.298	0.6093	0.9554
S.D.	0.4023	0.1867	0.07527	0.1914	0.4574	0.08486	0.1278	0.2493
Significantly different to ΔΔtlo	****	-	****	****	***	*	ns	ns
P-value	<0.0001	-	<0.0001	<0.0001	0.0005	0.0491	0.9995	0.6572
Significantly different to AHY940	-	****	ns	ns	**	***	****	****
P-value	-	<0.0001	0.9894	0.9997	0.0099	0.0001	<0.0001	<0.0001

Figure 5.20 Susceptibility of *TLO* reintroduction strains to oxidative stress

Minimum inhibitory concentration assays were performed by microtiter dilution to determine the IC_{50} of strains in oxidative stress inducing reagents. IC_{50} is the concentration at which growth is inhibited by 50%. (A) IC_{50} concentrations of each strain in hydrogen peroxide (H_2O_2), mean is indicated by the horizontal bar, with error bars representing standard deviation and symbols representing each of three replicates. (B) IC_{50} concentrations of strains in tBOOH. (C&D) Summary of data and statistical analysis (one-way ANOVA) of growth of strains in H_2O_2 and tBOOH respectively.

5.3.5.6 *Galleria mellonella* infection model

A *Galleria mellonella* infection model was used to determine how virulence was affected by the introduction of each *TLO* into the *ΔΔtlo* background. Kaplan-Meyer curves of the survival of the worms can be seen in Figure 5.21. The *ΔΔtlo* mutant strain was much less virulent than the WT. Reintroduction of *TLOα1* restored virulence in this model, and the strain expressing this gene from the *ENO1* promoter resulted in a higher level of virulence than the WT. Reintroduction of *TLOβ2* expressed under either promoter was also able to restore virulence to WT level. Reintroduction of *TLOγ11* regardless of the promoter, had no effect on virulence.



B

	AHY940	*ΔΔtlo*	P_{TET}TLOα1	P_{ENO}TLOα1	P_{TET}TLOβ2	P_{ENO}TLOβ2	P_{TET}TLOγ11	P_{ENO}TLOγ11
Significantly different to ΔΔtlo	***	-	****	****	***	****	ns	ns
P-value	0.0007	-	<0.0001	<0.0001	0.0002	<0.0001	0.0704	0.6003
Significantly different to AHY940	-	***	ns	*	ns	ns	*	**
P-value	-	0.0007	0.1127	0.0208	0.8367	0.1981	0.0267	0.0013

Figure 5.21 Survival of *Galleria mellonella* larvae after infection with *TLO* reintroduction strains

Each *Galleria mellonella* larva was infected with 10^6 cells in a 20µl volume. Inoculation with the same volume of PBS was used as a negative control. In total 20 worms were infected with each strain/PBS. (A) Kaplan-Meyer curve of survival of worms over 72 h after infection with *C. albicans* strains (B) Statistical analysis of results. Strains were compared to AHY940 and to the *ΔΔtlo* strain using a Log-rank (Mantel-Cox) test. Degrees of significance are indicated by asterisks, and P-values are given.

5.4 Discussion

5.4.1 Reintroduction of TLO *genes*

The data presented in Chapter 4 detailed the significant effects that the deletion of the *TLO* family had on the phenotype of *C. albicans*. Complementing deletions by reintroducing a wild type version of the mutated gene helps confirm the functions of deleted genes. It is impossible to reconstitute all 14 *TLO* genes, however, this chapter describes the results of reintroducing three clade representative *TLO* genes (*TLOα1, TLOβ2* and *TLOγ11*) into the *ΔΔtlo* mutant strain. Each *TLO* gene was successfully integrated into the mutant under the control of either a strong *ENO1* promoter, or the weaker *TET1* promoter. *TLO* genes under the *ENO1* promoter were incorporated into the non-telomeric *TLOα34* locus as introduction of a gene into a telomeric site may be difficult due to the heterochromatic nature of telomeric DNA, and the propensity for recombination events to occur at chromosome ends (Anderson *et al.*, 2015). The fact that genes at telomeres are subject to telomere positional effect (TPE) silencing could also confound results in these experiments (Anderson *et al.*, 2014). The *TLO* genes under the *TET1* promoter were incorporated into the *ADH1* locus due to the nature of the pNIM1 cassette used for reconstituting the mutation.

The genomics company Genewiz synthesised the 3xHA tagged *TLO* constructs under the control of the *ENO1* promoter with homology at each end for integration at the *TLOα34* locus. For unknown reasons they were unable to generate similar constructs where each *TLO* gene was under the control of the *TLOα1* promoter. The set of strains under the *TLOα1* promoter were supposed to represent a native level of expression, and attempts were made in the laboratory to clone these constructs in house, but since generation of these strains was unsuccessful, a new plan to generate a set of strains with the *TLO* genes under the *TET1* tetracycline inducible promoter was devised.

The introduction of each *TLO* cassette at the required locus was confirmed by routine PCR, Figure 5.3. Expression analysis was carried out using qRT-PCR to determine mRNA expression levels, Figure 5.4, and Western blotting, targeting the 3xHA tag, was used to confirm expression of protein from these constructs, Figure 5.5,. For the purposes of this work, the leaky expression from the tetracycline inducible promoter was sufficient to represent a relatively low level of *TLO* expression. Leaky expression from tetracycline

inducible promoter systems has been described in previous publications (Salucci *et al.*, 2002), and there was also concern that inclusion of tetracycline into phenotypic assays could affect phenotypes. Tetracycline has been reported to promote hyphal growth of *C. albicans*, and was also seen to have effects on growth rate in YNB and on biofilm formation (McCool *et al.*, 2008). mRNA and protein expression was detected for the *TLOα1* and *TLOβ2* strains, with a detectable difference evident in expression from the *ENO1* and *TET1* promoters.

In the case of both strains containing the reintroduced *TLOγ11* genes, while mRNA was clearly expressed from the reintroduced gene, no protein was detected by Western blotting under the culture conditions used. Western blots were performed with increasing concentrations of cell lysate, with various extraction methods and even with a mitochondrial enrichment (data not shown) to try to detect Tloγ11 protein. Data from Anderson *et al.* published in 2012 indicated that the gene products from γ clade *TLO* genes may localise in the mitochondria. These data also showed that increased concentrations of protein, up to ten-fold, were required to visualise a GFP tagged Tloγ5 via Western blot (Anderson *et al.*, 2012). It is possible that the translation of *TLOγ11* mRNA to protein is very tightly controlled in the *C. albicans* cell and that no matter the mRNA expression level, Tloγ11 protein levels remain low. Initially, it was hypothesised that the alternative splicing that γ clade *TLO* genes undergo could have effectively removed the 3xHA tag sequence from the mRNA thus rendering the protein undetectable (Anderson *et al.*, 2012). Interestingly, when the same *ΔΔtlo::*P_{ENO}*TLOγ11* construct was expressed in the WT AHY940 background, Tloγ11 protein was detected using mass spectrometry (performed by Dr Dean Frawley (a postdoc in the lab)) but not by Western blot. In addition, ChIP-seq experiments described in Chapter 6 showed that it was possible to pull down Tloγ11 protein using the same anti-HA antibody used in the Western blot analysis performed here. Therefore it appears that more sensitive detection methods are required to detect the Tloγ11 protein under routine culture conditions.

While the pNIM1 vector used to introduce the *TLO*s into the *ADH1* locus under the *TET1* promoter was not designed to be recyclable (Park and Morschhauser, 2005), the pJESS vector, a derivative of pSFS2a (Reuss *et al.*, 2004) was supposed to allow for recycling of the nourseothricin cassette from successful transformants via a flippase gene (*CaFLP*) under the control of a maltose inducible promoter (*MAL2*p) housed within the *SAT1* flipper segment. Many attempts were carried out with various maltose containing

media and various concentrations of maltose in order to try and recycle the nourseothricin resistance cassette. The original reference describing the *SAT1* cassette seems to indicate that the expression from the segment of the *MAL2* promoter included in the *SAT1* cassette was leaky, and that in many cases, no induction via maltose was necessary for precise excision of the cassette, and that growth in YEPD was sufficient for excision (Reuss *et al.*, 2004). Failure to excise the cassette in these experiments may be a consequence of the lack of *TLOs* in the deletion mutant. In *S. cerevisiae*, the *ScFLP1* gene, a homologue of the *CaFLP* gene used in the *SAT1* flipper, is located on the 2-µm plasmid, along with genes *REP1, REP2* and *RAF1*. In work performed by Ronne and Björklund *et al.* in 2004, it was seen that Srb10-dependent phosphorylation of the Med2 Mediator subunit was required for expression of these genes, including *FLP1,* from the 2-µm plasmid (Hallberg *et al.*, 2004). The Med2 subunit in *S. cerevisiae* is a homologue for the Tlo proteins in *C. albicans*, and the Srb10 protein of *S. cerevisiae* is a homologue of the Ssn3 protein in *C. albicans,* a subunit of the Cdk8 subunit of the Mediator complex along with Ssn8. In *C. albicans,* Tlos can be phosphorylated by Cdk8 kinase (Willger *et al.*, 2015). Therefore, underlying defects in these phosphorylation pathways caused by the complete loss of *TLO*s in the *ΔΔtlo* mutant strain could result in a failure of flippase activity and consequent inability to excise the cassettes.

5.4.2 Phenotypic analysis of reconstituted strains

A wide range of phenotypic tests were performed on the parent, deletion mutant and reconstituted strains in order to determine if there were differences in the ability of each *TLO* gene to restore the mutant phenotypes. A summary of the results of these tests can be seen in the heat map in Figure 5.22. From these data it is evident that *TLOα1,* expressed from both the *TET1* and *ENO1* promoters, was able to restore all of the phenotypes of the mutant so that it was practically indistinguishable from the parent strain. It should be noted that the AHY940 strain has been found to possess three copies of chromosome 5, therefore is not a true WT strain, however it is the parent strain for the *ΔΔtlo* mutant and all strains generated in this chapter. This trisomy is resolved in the *ΔΔtlo* mutant and all subsequent strains.

In this study, the reintroduction of *TLOα1* into the *ΔΔtlo* mutant was able to restore the morphology to a normal yeast blastospore morphology similar to that of the WT, Figure 5.6, where the mutant grew as pseudohyphae. Both of the *TLOα1* expressing strains were able to form true hyphae after induction with temperature shift and 10% FCS,

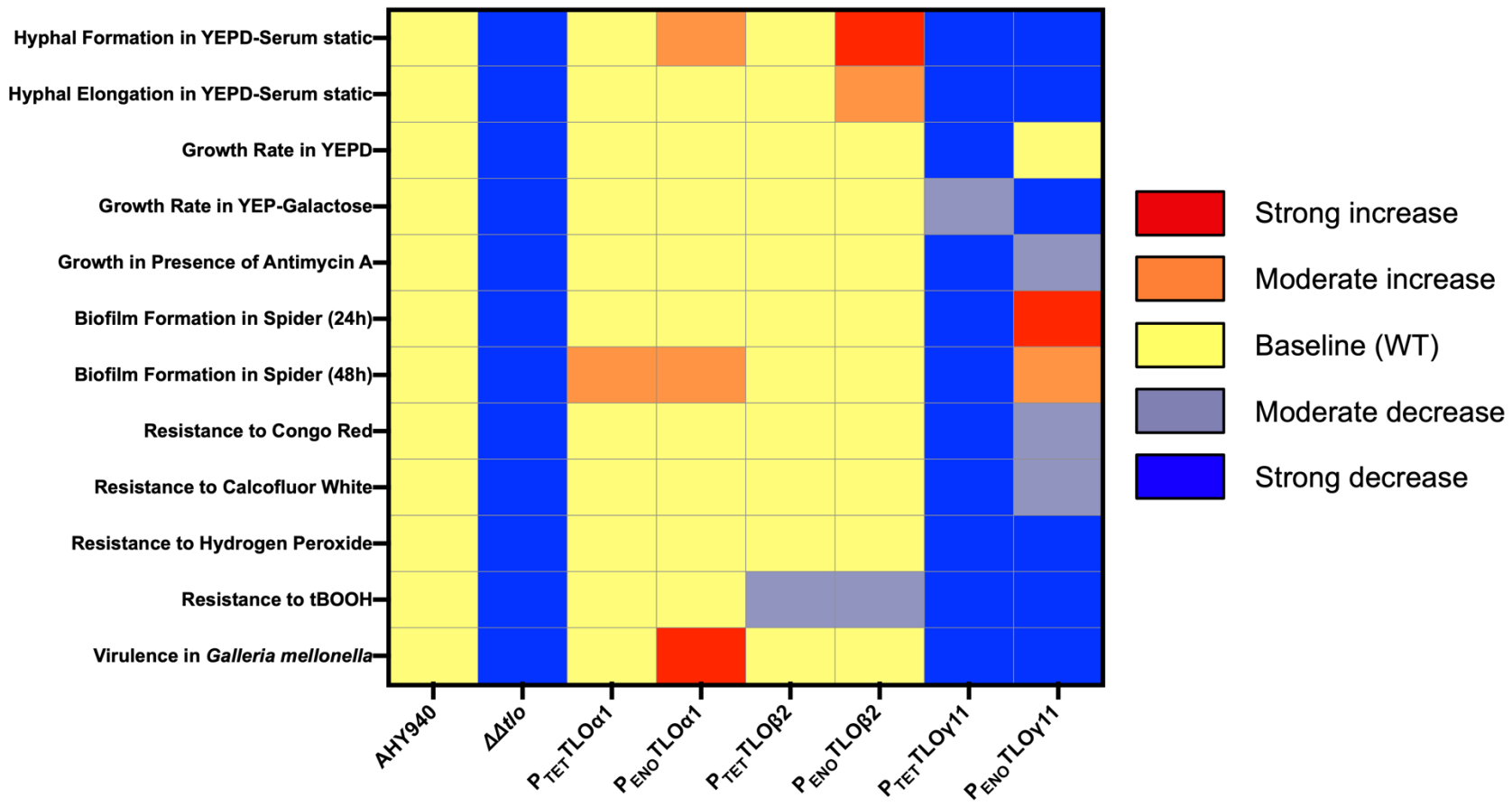


Figure 5.22 Heatmap of phenotype assay results in *C. albicans TLO* reintroduction strains

Phenotypes were graded compared to the AHY940 WT strain and the *ΔΔtlo* mutant strain response. Colour scale can be found on the right hand side.

Figure 5.7, while the *ΔΔtlo* mutant could not form true hyphae under any of the inducing conditions tested. Morphologies were determined based on the descriptions in (Sudbery *et al.*, 2004). In a static hyphal induction assay the *ΔΔtlo::*P_{ENO}*TLOα1* formed significantly more hyphae (measured as a percentage of cells displaying true hyphal morphology one hour post induction with temperature shift and 10% FCS) than the WT strain, Figure 5.10. The lengths of these hyphae remained similar to those of the WT strain. In the 2018 work by Dunn *et al.* it was shown that increasing the expression of *TLOα9* could increase the ability of the strain to filament in Spider media (Dunn *et al.*, 2018).

In a *C. dubliniensis Δtlo1/Δtlo2* null mutant, the introduction of *CaTLOα1* was able to increase growth rate in YEP-Galactose, and increase resistance to oxidative stress induced by H_2O_2 (Flanagan *et al.*, 2018). Similar results were obtained in this work, with the addition of *TLOα1* to the *ΔΔtlo* mutant *C. albicans* strain restoring growth rate in YEP-Galactose, Figure 5.14, and increased resistance to oxidative stress, Figure 5.20. Resistance to cell wall perturbing compounds, such as Congo Red and Calcofluor White, was also increased by the addition of *TLOα1* to the *TLO* null background *C. albicans,* Figure 5.19. Interestingly, when *CaTLOα1* was expressed in the *Δtlo1/Δtlo2 C. dubliniensis* strain it was not able to restore this phenotype, however, both *CaTLOα3* and *CaTLOα12* could restore resistance to Calcofluor White, while *CaTLOα3* could restore resistance to Congo Red (Flanagan *et al.*, 2018).

Work by Dunn *et al.* 2018 generated a set of strains where one copy of a selected *TLO* gene in a WT strain if *C. albicans* was put under the control of a doxycycline inducible promoter. In DOX^- conditions, the expression of the *TLO* was decreased, but not completely abolished due to expression from the other allele under the native promoter, and in DOX^+ conditions, the expression of the *TLO* was increased. While *TLOα1* was actually one of the only *TLO* genes that this group did not look at (the other was *TLOα10*), they did produce data on the other α clade *TLOs* (Dunn *et al.*, 2018). Much of their analysis into the *α* clade *TLO*s showed that increasing the expression of one α clade gene vs decreasing it caused an increase in phenotype, as they call it, for example, increasing the expression of *TLOα3* increased resistance to NaCl stress and stress induced by hydroxyurea compared to when expression of *TLOα3* was reduced.

Another interesting difference in phenotype seemingly caused by the increased expression of *TLOα1* was the increase in virulence in the *Galleria mellonella* model seen

in the *ΔΔtlo::P_{ENO}TLOα1* strain, Figure 5.21. While the *ΔΔtlo::P_{TET}TLOα1* strain displayed a similar level of virulence to the WT, which is more virulent than the *ΔΔtlo* mutant, the *ΔΔtlo::P_{ENO}TLOα1* strain was significantly more virulent than the parent strain, AHY940. In the Dunn *et al.* study, it was shown that increased expression of *TLOα34* led to an increase in virulence in the same infection model (Dunn *et al.*, 2018). A *Δtlo1/Δtlo2* strain of *C. dubliniensis* complemented with *CaTLOα1* displayed an increased level of virulence in the *G. mellonella* model, as did strains complemented with *CaTLOα3* or *CaTLOα9* (Flanagan *et al.*, 2018). Expression level of *CaTLOs* also proved to be important in the *G. mellonella* model of virulence in the work performed in Chapter 3, where an increased the level of expression of either *CaTLOα12, CaTLOβ2* or *CaTLOγ11* under the *ACT1* promoter in WT *C. dubliniensis* resulted in enhanced virulence in the model than the same genes being expressed in WT *C. dubliniensis* under their native promoter. Comparative transcriptomic analysis has previously shown that *TLOα1* is expressed at higher levels *in C. albicans* infecting *G. mellonella* compared to in the systemic mouse model, perhaps suggesting that increased expression of *TLOα1* may confer some form of selective advantage in this particular infection model (Amorim-Vaz *et al.*, 2015).

It is possible that the increased virulence observed in the *G. mellonella* model with the *ΔΔtlo::P_{ENO}TLOα1* reintroduction strain compared to *ΔΔtlo::P_{TET}TLOα1* could be Mediator independent. It has been shown that in WT *C. albicans* there is a large pool of free Tlo protein in addition to the Tlo that is associated with the Mediator complex (Zhang *et al.*, 2012). It may be the case that in the *ΔΔtlo::P_{TET}TLOα1* strain all the Tloα1 protein produced is incorporated into the Mediator complex and that there is no free pool of Tloα1 protein. The *ΔΔtlo::P_{ENO}TLOα1* strain could be producing enough protein to incorporate into Mediator and to also be present as free Tlo. Tlo proteins in *C. albicans* not only contain a conserved Med2 domain for integration into the Mediator complex, but also a transcriptional activation domain (TAD). It is believed that free Tlos can enact regulatory functions via these TADs, and it has been shown that in *C. dubliniensis,* a *CdTLO2* TAD segment fused to a nuclear localisation sequence was enough to drive filamentation phenotypes seen when *CdTLO2* itself was overexpressed (Liu *et al.*, 2016). It seems likely that the TAD of *TLOα1* plays a role in the increased virulence in the *G. mellonella* model. Expressing a *TLOα1* TAD with a nuclear localisation sequence in the *ΔΔtlo* background could decipher if this phenotype relies on free Tlo and TAD activity rather than Mediator associated Tlo. Proposals to use a *ΔΔtlo/ΔΔmed3* background to examine the role of reintroduced *TLOs* may be complicated by the fact that Tlo proteins require Med3 and Med15 for recruitment into the nucleus (Liu *et al.*, 2016). If the Tlos

were not recruited into the nucleus they would not be able to interact with the DNA in either state. If performed, these experiments could uncover another role for Tlo that is neither Mediator associated or TAD activity dependent.

The reintroduction of *TLOα1* into the *C. albicans ΔΔtlo* mutant can restore the ability of the strains to form biofilms in Spider medium on plastic surfaces, Figure 5.18. In transcriptomic analysis by Uppuluri *et al.* in 2018 it was shown that a subset of *TLO* genes are upregulated in biofilm-associated cells versus planktonic cells or dispersed cells. In this subset, the only α clade *TLO* that was upregulated was *TLOα1*, all the other upregulated *TLO*s were from the β or γ clade (Uppuluri *et al.*, 2018).

In the *C. dubliniensis* work performed in Chapter 3, the introduction of another α clade *TLO, TLOα12* increased the ability of WT *C. dubliniensis* to form biofilms, Figure 3.8. And the work by Dunn *et al.* in 2018 links the increase in expression of *TLOα3* and *TLOα34* versus the decrease in their expression to being associated with an increase in biofilm formation (Dunn *et al.*, 2018).

Dunn *et al.* (2018) performed principal component analysis on phenotypic data from *TLO* up/down regulation strains to determine if there was any correlation between *TLO* clade and the phenotypes they saw affected in their work. Results from this were not clear cut, but there was some clustering of *TLO*s within the same clade suggesting that these proteins may have clade and protein specific functions(Dunn *et al.*, 2018).

In this analysis the *TLOα1* gene appeared to be able to restore all phenotypes tested which could hint that in a background where there are no other *TLOs* present to play the roles they are adapted to, *TLOα1* can perform these roles. The fact that the ancestral *TLO* gene is thought to have been most similar to the α clade *TLOs* could also be a factor in the ability of the *TLOα1* to take up so many roles in a cell lacking all other *TLO*s (Anderson *et al.*, 2012). As in most species which possess a single *MED2* gene, the ancestral *TLO* likely had a role in many cellular pathways. Subsequent evolution and expansion of the *TLO* family in *C. albicans* may have resulted in the role of each *TLO* gene to become more specialised in the subset of genes that they help regulate. As *TLOα1* was the only α clade gene examined in this study, it would be interesting to see if other members of this clade had the same ability as *TLOα1* to restore such a broad range of phenotypes.

In comparison with *TLOα1*, introduction of the *TLOβ2* gene into the *ΔΔtlo* null background had a more specific effect on the morphology of the cell. The wrinkled morphology of the strains on YEPD agar, and the fact that many cells displayed a truly hyphal morphology in the absence of hypha-inducing conditions, Figure 5.6, suggests that there may be a role for *TLOβ2* in the hyphal development pathway of *C. albicans*. Overexpression of *TLOβ2* from the *ENO1* promoter also increased the amount of hyphal cells seen in non-inducing conditions, and when hyphal morphology was induced, this strain formed significantly more hyphae than the WT or the *ΔΔtlo::*P_{TET}*TLOβ2*, and these hyphae were longer, Figure 5.10. This indicates a gene dosage dependent role for *TLOβ2* in the hyphal development pathway.

Similar phenotypes were noted in Chapter 3, where *CaTLOβ2* was heterologously expressed at high and low levels in *C. dubliniensis* strains. These strains displayed the wrinkled colony morphology on YEPD agar, true hyphal cells were observed under the microscope, and there appeared to be a correlation between a higher level of expression of *CaTLOβ2* and the amount of wrinkling and hyphal growth, Figure 3.3.

In work by Dunn *et al.* in 2018, decreasing the expression of *TLOβ2* did not result in any defects in filamentation in *C. albicans* strains. This could indicate that even a small amount of *TLOβ2* expression is enough to induce the hyphal morphology (Dunn *et al.*, 2018). It could also be the case that, as described and displayed above, other *TLO*s are able to perform the role of *TLOβ2* under certain conditions when *TLOβ2* is absent. In the work performed by Flanagan *et al.* in 2018 on a *Δtlo1/Δtlo2* strain of *C. dubliniensis, CaTLOβ2* was not able to be expressed in the background, but expression of various α clade *TLOs* or *TLOγ11* could restore the ability of strains to form true hyphae in inducing conditions (Flanagan *et al.*, 2018).

When the transcriptomes of *C. albicans* biofilm and dispersed cells were compared previously, the expression of *TLOβ2* was noted as being one of the most differentially expressed genes as it was expressed seven-fold higher in biofilm associated cells (Uppuluri *et al.*, 2018). When *TLOβ2* was specifically deleted there was up to a 60% decrease in biofilm formation. The suggestion that deletion of one single *TLO* gene had such a significant phenotype is surprising and does not correlate with the data from this study, where a strain only expressing *TLOα1* or *TLOγ11* under the *ENO1* promoter could produce levels of biofilm comparable to, or in excess of, that of the WT strain. There

were, however, differences in the biofilm assays used in both studies, e.g. YNB medium used by Uppuluri *et al.* and Spider medium used in this study.

Reintroduction of *TLOβ2* to the *ΔΔtlo* background was able to complement many phenotypes, however one exception was the response to oxidative stress induced by tBOOH, Figure 5.20. While the reintroduction of *TLOα1* (under either promoter) to the *ΔΔtlo* background was able to restore tolerance to oxidative stress induced by tBOOH to WT levels, the reintroduction of *TLOβ2* resulted in a lower level of tolerance compared to the WT and *TLOα1* reintroduction strains. tBOOH, *tert*-butyl hydroperoxide, is an oxidative stress generating agent that is thought to be more stable than H_2O_2 and is often used as a oxidative stress agent in sensitivity assays (Kaloriti *et al.*, 2012). It is thought that the damage caused by tBOOH is similar to the damage caused when cells are phagocytosed and phagocytes generate reactive oxygen species (ROS) to fight off invading pathogens. The ROS damage the lipids in the cell walls via lipid peroxidation (Fekete *et al.*, 2007). It has been noted that strains that have been selected based on their ability to tolerate oxidative stress induced by tBOOH generally correlated with the strains losing the ability to form hyphae and displaying decreased virulence (Fekete *et al.*, 2007). It could be the case that the propensity for hyphal formation in the *TLOβ2* reintroduction strains may be a cause of their increased susceptibility compared to the WT and *TLOα1* reintroduction strains. An RNA-seq based approach could determine if it is in fact the hyphal morphology of the cells that is contributing to the increased susceptibility of the *TLOβ2* reintroduction strains.

The survival curves generated from the *G. mellonella* infection models for both of the *TLOβ2* reintroduction strains are similar to the WT, Figure 5.21. The immune response of *G. mellonella* to infection with bacteria or fungi is very similar to the human innate immune response. This response is mainly mediated by haemocytes, which are cells that perform similar functions to human phagocytes. *Galleria mellonella* haemocytes have been shown to produce ROS, including nitric oxide, superoxide and H_2O_2 (Pereira *et al.*, 2018). So although the *TLOβ2* reintroduction strains are more sensitive than the WT to tBOOH induced oxidative stress in an *in vitro* assay, this does not appear to impact virulence in the *in vivo* assay.

The ability to switch between the yeast and hyphal forms are thought to contribute to the increased virulence of *C. albicans* in comparison with other *Candida* species and it has been shown that cells locked in either form cannot cause blood stream infections in

a mouse model. For example a *ΔΔnrg1* mutant strain that cannot form true hyphae is avirulent in a mouse model (Murad *et al.*, 2001). Recent work that shows that a *ΔΔeed1* mutant strain that cannot form hyphae, but which displays increased proliferation rates, can cause infection when injected directly into the bloodstream of mice. However, it is still believed that the ability to form hyphae is essential for penetration of the epithelium in other infection models (Dunker *et al.*, 2021). The *TLOβ2* reintroduction strains do not uniformly form hyphae; many of the cells remain as yeast in non-inducing conditions, but the fact that some of the cells grow as hyphae in YEPD at 37 °C suggests that that the control of the pathways required for hypha formation are disrupted in these strains. This is supported by the increased formation of hyphae seen in the *TLOβ2* reintroduction strains compared to the WT in YEPD with 10% FCS. However, this enhanced ability to produce hyphae does not lead to a commensurate increase in virulence in the *G. mellonella* model, perhaps due to other phenotypes, such as the lower tBOOH tolerance described above.

In *TLO* depleted strains that were generated in the process of obtaining a *TLO* null mutant strain of *C. albicans*, a strain that had all *TLO*s except for *TLOβ2* and *TLOα10* deleted (*ΔΔtlo*::*TLOβ2,α10*) displayed phenotypes that align with what was seen in the *TLOα1* and *TLOβ2* reintroduction strains in this chapter. Phenotypic tests were performed on this, the results of which can be found in Appendix 1. Briefly, the *ΔΔtlo*::*TLOβ2,α10* strain was able to perform similarly to the parent (SC5314 in this case) in many of the tests performed, such as growth rate analysis in YEPD and YEP-Galactose, chlamydospore formation, resistance to cell wall perturbing compounds, resistance to hydrogen peroxide, and virulence in a *G. mellonella* model. The *TLOα1* and *TLOβ2* expressing strains described in this chapter also behaved similarly to the parent AHY940 in these tests. Where the *ΔΔtlo*::*TLOβ2,α10* strain differed from the WT was in ability to form biofilms on plastic surfaces in Spider medium, and in ability to tolerate oxidative stress. The *TLO* depleted strain formed less biofilm than the parent, and was more sensitive oxidative stress induced by tBOOH. There is also some observational evidence that the lengths of hyphae formed by this strain are longer than that of the parent. In the case of biofilm formation, it is possible that the difference in parental strain could be behind the differences in phenotype in these strains. The level of biofilm formed by the AHY940 WT strain is much lower compared to that of the SC5314 WT. As described earlier, the AHY940 strain lacks one copy of the *LEU2* gene. The Leu2 protein has previously been extracted from *C. albicans* biofilms, and has been identified as a surface-associated protein in biofilm cells (Martinez *et al.*, 2016). The loss of one copy

of *LEU2* could be partly responsible for the AHY940 strain, and subsequent strains constructed from this parent, forming less biofilm than the SC5314 strain and strains constructed from this parent. The trisomy of chromosome 5 in the AHY940 strain could also be responsible for variations in phenotype compared to SC5314.

The sensitivity of the *ΔΔtlo*::*TLOβ2,α10* strain to tBOOH can also be compared to the results seen in the *TLOα1* and *TLOβ2* reintroduction strains in this chapter. While the *TLOα1* reintroduction strains have a similar resistance level to tBOOH as the parent strain, the *TLOβ2* reintroduction strains have an intermediate resistance level in between that of the parent strain, AHY940, and that of the *ΔΔtlo* and *TLOγ11* reintroduction strains. This is similar to what is seen in the *ΔΔtlo*::*TLOβ2,α10* strain, which displays a resistance level intermediate between the parent strain, SC5314, and the *TLO* depleted strain with only *TLOγ5* remaining. It was speculated above that it could be the increased propensity for hyphal formation that is responsible for this increased susceptibility of the *TLOβ2* reintroduction strains to tBOOH, and while no formal quantification of hyphal formation was performed on the *ΔΔtlo*::*TLOβ2,α10* strain, observational difference in the lengths of hyphae was noted, perhaps indicating that this *ΔΔtlo*::*TLOβ2,α10* strain could also be more primed for forming hyphae. It appears that *TLOβ2* plays an important role in hyphal development in the cell, and in the absence of other Tlo proteins, the hyphal morphology appears to not be completely repressed in conditions where the WT exists in the yeast form alone. It could be that the depletion of all other *TLO*s in the *ΔΔtlo*::*TLOβ2,α10* strain results in a strain more primed for hyphal development due to an imbalance of Tloβ2 protein in the cell.

The reintroduction of *TLOγ11* into the *ΔΔtlo* background still remains somewhat enigmatic. mRNA was readily detected from both promoters, however, no protein was detectable using Western blotting, suggesting that the protein is not translated or is expressed at levels below the threshold of detection using this method.

Work by Anderson 2012 showed how the different clades of *TLO* are differentially spliced and also how they localise within the cell. They were unable to confirm that the *TLOγ11* transcript undergoes alternative splicing, but they did state that the *TLOγ11* sequence does not contain a nuclear localisation sequence (Anderson *et al.*, 2012). The genome assembly published by van het Hoog *et al.* in 2007 predicted that all seven of the γ clade *TLO*s were able to produce two distinct RNA isoforms, but only four of these; *TLOγ5, TLOγ7, TLOγ13* and *TLOγ16*, have been confirmed (van het Hoog *et al.*, 2007,

Anderson *et al.*, 2012). The *TLOγ13* and *TLOγ16* spliced transcripts were found to be unlike those for any other *TLO* gene characterised, as these two transcripts did not have a 3´ UTR. They predicted that this would negatively impact transcript stability and make translation less efficient (Anderson *et al.*, 2012). Although they did not refer to *TLOγ11,* there may also be an issue with efficiency of translation of this gene. Interestingly, Tloγ11 protein has been detected by mass spectrometry, (Dr Dean Frawley, personal communication) and data from the ChIP-seq analysis described in the next chapter suggests that this protein is expressed.

Despite the fact that there were difficulties in detecting Tloγ11 protein (at least in the growth conditions used) expression of this gene, particularly from the *ENO1* promoter, did complement some of the phenotypes of the *ΔΔtlo* null mutant.

Resistance to oxidative stress, Figure 5.20, and virulence in the *Galleria mellonella* model, Figure 5.21, were not affected by the reintroduction of *TLOγ11* and these strains performed similarly to the *ΔΔtlo* mutant strain. This suggests that *TLOγ11* may not be involved in these phenotypes. Similarly, the reintroduction of *TLOγ11* did not restore the morphological defects associated with the mutant, i.e. pseudohyphal growth and inability to form hyphae, Figures 5.5 & 5.6, suggesting that *TLOγ11* is not involved in morphology related or hyphal development pathways. One outlier to the trend was the fact that the *ΔΔtlo::*P_{TET}*TLOγ11* strain was actually able to grow faster than the *ΔΔtlo::*P_{ENO}*TLOγ11* strain and the *ΔΔtlo* mutant in YEP-Galactose, Figure 5.14. This growth rate difference does not appear to be morphology dependent as the strain displays the same pseudohyphal morphology as the *ΔΔtlo::*P_{ENO}*TLOγ11* strain and the *ΔΔtlo* mutant when grown in YEP-Galactose. Pseudohyphal morphology does not always result in slowed growth rate, as seen in a *ΔΔtup1* mutant strain of *C. albicans* that grows as pseudohyphae but has no difference in exponential growth rate from the WT (Braun and Johnson, 1997). Contrary to this, the *ΔΔtlo::*P_{ENO}*TLOγ11* strain grew faster than the *ΔΔtlo::*P_{TET}*TLOγ11* strain and the *ΔΔtlo* mutant in YEPD, Figure 5.14. While this result is interesting, it was not investigated further as a part of this work. More investigation would be required in order to determine what effect the low level of expression of *TLOγ11* is having on this growth rate that the higher level of expression is not.

The *ΔΔtlo::*P_{ENO}*TLOγ11* strain is more resistant to cell wall stress than the *ΔΔtlo::*P_{TET}*TLOγ11* strain and the *ΔΔtlo* mutant, but it is still not as resistant as the WT or the *TLOα1* or *TLOβ2* reintroduction strains, Figure 5.19. Further analysis would be

required to understand what is going on in these cells, but a possible candidate for investigation could be the Mkc1 protein, a stress-response activated MAP kinase involved in responding to various cell wall stresses, loss of *MKC1* decreases resistance to CFW and CR (Navarro-Garcia *et al.*, 2005). MAP kinases play roles in signal transduction and the regulation cellular pathways in *C. albicans* including stress resistance, biofilm formation, hyphal formation. They themselves require phosphorylation to active, and they can then go on to phosphorylate other proteins (Willger *et al.*, 2015).

Increased phosphorylation of Mkc1 is seen after introduction of cell wall stress with CFW or CR. This increase in phosphorylation of Mkc1 is also seen when cells are exposed to oxidative stress by H_2O_2, but this phosphorylation has not been found to be essential for cell survival in oxidative stress conditions. Other MAP kinases such as Hog1 and Pkc1 have also been implicated in the stress response pathways (Navarro-Garcia *et al.*, 2005). A difference in the ability of the *ΔΔtlo::*P_{ENO}*TLOγ11* strain to phosphorylate and activate these MAP kinases could explain the differences in the ability to resist cell wall stress, and the fact that there is no difference in the resistance levels to oxidative stress seems to fit this narrative. The relationship between *TLO*s and phosphorylation has been highlighted previously in this discussion, where it was speculated that defective phosphorylation pathways could contribute to the failure of the excision of the *SAT1* flipper in these strains. There are almost 16,000 estimated phosphorylation sites in the *C. albicans* proteome located on 2,896 different proteins (Willger *et al.*, 2015), and mis-regulation of the phosphorylation pathways within the cells could have detrimental effects on a cell's fitness and virulence. It is very likely that the phosphorylation pathways in the *ΔΔtlo* mutant are impacted in some way. Investigation into the phosphoproteomes of these strains could uncover any deficiencies in phosphorylation in these strains.

The most apparent difference between the *ΔΔtlo::*P_{ENO}*TLOγ11* strain and all the others is the amount of biofilm formed by this strain after 24 h in Spider medium, Figure 5.18. The level of biofilm measured in this assay was much higher than even the WT strain. In Chapter 3, expression of *CaTLOγ11* under either a native promoter or an *ACT1* overexpressing promoter in a WT *C. dubliniensis* strain also increased the ability of these strains to form biofilms compared with the WT strain.

Data from Uppuluri *et al.* in 2018 showed that eight *TLO* genes were upregulated in biofilm cells in *C. albicans*, and six of the eight upregulated *TLO*s were members of the γ clade, including *TLOγ11* (Uppuluri *et al.*, 2018). While none of the γ clade *TLO*s were the most highly upregulated in biofilm formation, it is interesting to note that this seems to be a phenotype that involves the expression of this clade.

Work performed by Flanagan *et al.* in 2018 found that expression of *CaTLOγ11* in a *Δtlo1/Δtlo2 C. dubliniensis* mutant strain did not affect this mutant's ability to form biofilm (Flanagan *et al.*, 2018). However, in this they also found that the *Δtlo1/Δtlo2* mutant strain of *C. dubliniensis* produced more biofilm that the WT Wü284 strain. These results conflict with the results in Chapter 3 of this work, where the opposite was found true, the *Δtlo1/Δtlo2* mutant strain of *C. dubliniensis* formed less biofilm than WT when using a crystal violet staining assay, although differences in experimental methods could account for these differences.

A *TLO* depleted strain of *C. albicans* (SC5314) with all *TLO*s deleted apart from *TLOγ5* (*ΔΔtlo::TLOγ5*) displayed similar phenotypes to those seen in the *TLOγ11* reintroduction strains analysed here, specifically the *ΔΔtlo::*P_{TET}*TLOγ11* strain. Results from the phenotypic analysis of this strain can be found in Appendix 1. Like the *ΔΔtlo* mutant and the *TLOγ11* reintroduction strains, this *ΔΔtlo::TLOγ5* had a pseudohyphal morphology, and could not form true hyphae in inducing conditions, the strain could not form chlamydospores, nor could it grow well in the presence of antimycin A. The *ΔΔtlo::TLOγ5* strain was much more sensitive to oxidative stress induced by either H_2O_2 or tBOOH, and was significantly less virulent in the *G. mellonella* model, again similar to the *ΔΔtlo* mutant and the *TLOγ11* reintroduction strains. The *ΔΔtlo::TLOγ5* strain formed significantly less biofilm than the parent strain, similar to what was observed in the *ΔΔtlo* mutant and the *ΔΔtlo::*P_{TET}*TLOγ11* strain. This in contrast to the *ΔΔtlo::*P_{ENO}*TLOγ11* strain which forms a large amount of biofilm in these conditions. This difference could be due to the high level of expression of *TLOγ11* from the *ENO1* promoter, compared to the native expression level from the *TLOγ5* promoter, as it has been documented that the γ-clade *TLOs* are expressed at a much lower level than the other two clades in the WT cell (Anderson *et al.*, 2012). The *ΔΔtlo::TLOγ5* strain was not significantly slower growing in YEPD than the parent strain, but was slightly slower growing in YEP-Galactose (however, this result was not found to be statistically significant). These phenotypes also correlate somewhat with the results of growth rate analysis performed on the *TLOγ11* reintroduction strains, where the *ΔΔtlo::*P_{ENO}*TLOγ11* strain grows as well

as the parent in YEPD, and where the *ΔΔtlo::P_{TET}TLOγ11* grows objectively slower than the parent strain, but not statistically significantly.

When *CaTLOγ11* was expressed in a *Δtlo1/Δtlo2* mutant strain of *C. dubliniensis* it was able to restore multiple phenotypes, including some that weren't affected when the same gene was expressed in the *C albicans* mutant. In contrast to the *C. albicans* experiments, in the *C. dubliniensis* background where expression of *CaTLOγ11* was controlled by its native promoter, it was found that chlamydospore and hypha formation (in water supplemented with 10% FCS) was restored. Similarly, tolerance of oxidative stress, was restored in the *C. dubliniensis*, but not the *C. albicans*, mutants (Flanagan *et al.*, 2018). These discrepancies may be a reflection of the different *TLO* landscape in the two species, but suggests that *CaTLOγ11* has the potential to regulate a wide range of phenotypes. Unfortunately, no experiments were performed to investigate if Tlo protein expression was detectable in either the WT or mutant *C. dubliniensis* backgrounds.

In the experiments described in this chapter, the effects of a limited number of representative *TLO* genes were examined in isolation. Clearly, in order to obtain a definitive understanding of the functionality of each gene and the relevance of the clades a comprehensive analysis of the gene family would require that each gene was reconstituted either alone or in conjunction with other members of the clade. Another avenue for future investigation would be to introduce multiple *TLO*s into these background together to determine if *TLO*s act together or antagonistically. One hypothesis to help explain the increased virulence of *C. albicans* compared to other species, is that the expansion of the *TLO* repertoire allows for cells to be more phenotypically plastic (Liu *et al.*, 2016). The incorporation of different sets of *TLO* genes into a *ΔΔtlo* background would allow *TLO* interaction and competition to be studied. While the data presented here suggest that there are no deficiencies in the cell expressing only *TLOα1,* further in depth phenotypic testing could uncover phenotypes that require the presence of multiple *TLO*s. The relationships between Tlo proteins and Mediator could also be investigated further, it could be the case that inclusion of certain Tlos is favoured in certain conditions, and that another Tlo confers more of an advantage in the free state.

Expression of the *TLO*s in WT *C. albicans* is not uniform and may vary on a cell to cell level, this is related to TAGEN (described more in depth in Section 1.5.3). When selective

pressure is applied, the varied expression of Tlos in different cells may facilitate quicker adaption to changing environmental conditions (Anderson *et al.*, 2014). In the experiments described in this chapter, by expressing the *TLO* constructs from regions not affected by TAGEN (e.g. not at the telomeres), the effects seen may be somewhat artificial. These experiments still explore the roles of specific *TLO*s in pathways and phenotypes, but under unrealistic expression conditions. Introducing the genes at their native loci, under native promoters, while possibly more complicated due to the telomeric nature of the genes, could provide a more true to life model for examining the roles of the *TLOs.* It may also be the case that there is less silencing at the telomeres in the *ΔΔtlo* strain or strains with fewer *TLOs*, as Tlo integration into the Mediator is believed to be a part of the silencing machinery through the Mediator's interaction with Sir2, the histone deacetylase that is responsible for much of the telomere associated silencing seen in *C. albicans* (Anderson *et al.*, 2014).

The work described in this chapter revealed that many phenotypes that were affected by the complete loss of *TLO*s in *C. albicans* could be restored by reintegration of a single *TLO*. Reintroduction of *TLOα1* was able to complement all phenotypes tested, possibly reflecting its relationship to the ancestral *TLO*. There were some examples of *TLO*s with very specific effects (e.g. the effect of *TLOβ2* on morphology), while *TLOγ11* clearly differs in its ability to restore the mutant phenotypes. In order to further investigate if specific Tlo proteins play a role in controlling different subsets of genes, RNA-sequencing experiments were designed to identify the transcriptomes of cells grown to mid-exponential phase in YEPD and also in the presence of tBOOH (e.g. reflective of oxidative stress). The latter conditions were chosen because each of the three TLO genes tested restored the phenotype of the deletion mutant strain to a different extent. ChIP-sequencing experiments were also completed to determine if the three Tlos examined interacted with similar or different areas of the genome. The results of these experiments are described in the next chapter.

Chapter 6

Transcriptomic analysis of *TLO* mutants in nutrient rich conditions and under oxidative stress

6.1 Introduction

RNA-sequencing is an offshoot of NGS where the transcriptome is sequenced rather than the genome. This process relies on the isolation of mRNA from a culture, conversion of this mRNA into cDNA and subsequent sequencing. RNA-sequencing is mainly used to examine differential expression of genes, be that between the same strain in different conditions, or between strains/mutants in the same conditions. Before RNA-seq technology was readily available, microarrays were the main tool used for determining differential gene expression between samples, but RNA-seq is more sensitive and allows for better measurement of differential expression analysis (Rao *et al.*, 2019).

RNA-seq analysis usually involves the alignment of generated sequencing reads to a reference genome, and then quantification of the coverage at each gene, which is related to the relative expression of that gene. Comparison between samples involves the use of software to compare the relative expression of the genes in each condition or strain, which then gives a fold change value representing the change in expression level of the gene in one sample versus the other. Genes that are found to be up- or downregulated can then be investigated further, for example though Gene Ontology (GO) analysis, which identifies the functions of genes, or through Gene Set Enrichment Analysis (GSEA) which determines if specific sets of genes are statistically different between two samples.

In *Candida albicans*, where having a readily adaptable transcriptome is key for virulence, RNA-sequencing has been used to examine transcriptomes under a wide variety of environmental conditions, specifically those related to virulence of the pathogen. Work published by Bruno *et al.* in 2010, examined the transcriptome of *C. albicans* in nine different conditions, including in YEPD + 10% FCS (known to induce hyphal formation), in a variety of pH conditions, in oxidative stress conditions and under cell wall damaging conditions. These experiments, which were performed relatively early in the RNA-seq timeline. They identified some novel transcripts which displayed altered expression in different conditions, they identified possible roles for many genes that were uncharacterised at the time, and they also identified additional roles for previously described genes (Bruno *et al.*, 2010). For example, they identified that the uncharacterised transcript C1_13100W_A (orf19.4936.1) was significantly upregulated in the presence of serum than in YEPD, and they proposed a role for this transcript in

the response to serum. They also uncovered novel roles for genes involved in arginine biosynthesis (*ARG1, ARG3, ARG4, ARG5, ARG6* and *ARG8*) and lysine biosynthesis (*LYS1, LYS2* and *LYS9*) in the oxidative stress response, where these genes were significantly upregulated in the presence of oxidative stress (Bruno *et al.*, 2010). RNA-sequencing by Nobile *et al.* published in 2012, described the transcriptional network governing biofilm formation in *C. albicans* and highlighted a set of six main regulators that control this phenotype (Nobile *et al.*, 2012). Another example of RNA-seq being used to study *Candida* spp*.* is work by Palige *et al.* published in 2013, who investigated the transcriptomes involved in chlamydospore formation in *C. albicans* and *C. dubliniensis*. This work compared chlamydospore forming WT strains of *C. albicans* to a *ΔΔnrg1* mutant which lacks the hyphal repressor gene and is constitutively pseudohyphal but still forms chlamydospores. A *C. dubliniensis* WT strain was also compared to the *C. albicans* WT, and the intersection of these two comparisons revealed a set of 25 genes specifically upregulated in chlamydospore formation and eight genes that are downregulated, with the group dubbing these putative chlamydospore specific genes (Palige *et al.*, 2013).

In an early study to investigate the role of the *TLO* genes in *C. dubliniensis,* the transcriptome of a *ΔCdtlo1/ΔCdtlo2* mutant grown to mid-exponential phase in nutrient rich YEPD conditions was compared to that of the WT *C. dubliniensis* strain using microarrays to determine which genes were impacted by the absence of Tlos (Haran *et al.*, 2014). Analysis of the array data showed that the transcriptome of the null mutant resembled that of a nutrient starvation response. There was downregulation of genes involved in glycolysis, such as *TYE7* and *ENO1,* and genes involved in amino acid biosynthesis, such as *SAM2* and *MET1*. There was upregulation of genes that were involved in alternative carbon and nitrogen metabolism, such as *NAG* genes for the metabolism of *N*-acetyl-glucosamine and *PUT1* and *PUT2* for metabolism of amino acids. Transcriptome analysis of this *Δtlo1/Δtlo2* mutant under hyphal inducing conditions, YEPD + 10% FCS, found that this strain did not express hyphal-specific genes, such as *UME6*, *EFG1*, *RAS1* and *RIM101*, at the same level as the WT. This result correlated with the inability of the mutant strain to form true hyphae under these conditions (Haran *et al.*, 2014).

6.1.1 Aims of this chapter

The main aims of this chapter were to use an RNA-Seq analysis to further investigate the role of Tlos in *C. albicans*, using strains described earlier in this work (i.e. the *ΔΔtlo* mutant, the strains where select *TLO*s were reintroduced under the control of the *TET1* promoter, and the AHY940 WT parent). RNA-seq analysis was carried out on cells grown in YEPD at 37 °C to mid-exponential phase and in the presence of the oxidative stress inducing compound tBOOH. The latter growth conditions were selected as it was observed in Chapter 5 that the WT and *ΔΔtlo* strains differ in their ability to tolerate oxidative stress. Similarly, differences in tolerance to tBOOH were observed between the *TLOα1*, *TLOβ2* and *TLOγ11* reintroduction strains, allowing us to hypothesise that comparative transcriptomic analysis might reveal differences in the activation of oxidative stress response pathways by specific *TLO*s.

6.2 Materials and Methods

Unless otherwise stated, the strains used in this chapter are the AHY940 parent strain, the *ΔΔtlo* mutant, and the *TLO* reintroduction strains, which consist of *TLOα1, TLOβ2* and *TLOγ11* expressed under the *TET1* promoter in the *ΔΔtlo* mutant background, see Table 2.2 for more detail.

6.2.1 Growth conditions

Cultures were grown overnight at 37 °C at 250 rpm in a shaking incubator. The next morning, the OD_{600} of the cultures was determined, and strains were equalised to an OD_{600} of 0.2 in 50 ml YEPD in an Erlenmeyer flask. Flasks were re-incubated at 37 °C at 250 rpm until OD_{600} reached approx. 0.8, indicating the cells were in exponential growth phase, after approx. 3-4 h. At this point, 25 ml of culture was removed for RNA extraction. In order to investigate transcriptomic responses to oxidative stress, tBOOH was added to the remaining 25 ml of culture to a concentration of 0.25 mM. This was then returned to the incubator for 30 min, after which the final 25 ml of culture was removed for RNA extraction.

6.2.2 RNA extraction

Cultures for RNA extraction were centrifuged at 2000 rpm for 5 min, the supernatants were removed, and the cell pellets were resuspended in 10 ml 1X PBS and centrifuged again as before. After centrifugation, 1 ml of supernatant was reserved, the rest was discarded. The cell pellets were then resuspended in the 1 ml of reserved supernatant. Cell suspensions were snap frozen in liquid nitrogen, by the addition of the cell suspension dropwise to a 50 ml tube containing around 35 ml of liquid nitrogen. Once the entire cell suspension was frozen, these pellets could be stored at -80 °C. To ensure integrity of the RNA, pellets were stored for as short a time as possible before extraction.

The RNeasy Mini Kit (QIAGEN) was used to perform RNA extractions. Cell pellets were removed from the freezer and kept on ice to prevent any melting. Pellets were added to a PTFE shaking flask with a tungsten carbide grinding ball, then processed in a Mikro-Dismembrator (Sartorius). Two rounds of disruption were performed sequentially, both for 1 min at 3000 rpm, with vigorous tapping of the shaking flask against a hard

surface in between. After processing, the flasks were opened and the grinding ball removed, the resulting powder was resuspended as per instructions of the RNeasy kit (Qiagen), and extraction proceeded as per the manufacturer's instructions. Final elution of RNA was performed in 30 µl water. The RNA samples were then treated with Turbo DNase (Thermo Fisher) to digest any DNA that may be present in the samples, as per the manufacturer's instructions.

At this point, the concentration and quality of the RNA was measured using a Nanodrop 2000c (Thermo Fisher). For sequencing purposes, a 260/280 quality ratio of between 1.8 - 2.2 was required, and a 260/230 ratio of ≥ 1.8. After initial extraction of RNA with the RNeasy kit it was found that in the majority of cases the 260/230 ratio was not adequate for sequencing. In these cases, the RNA sample was made up to 600 µl with RPE buffer from the RNeasy kit, and passed through a fresh column, with one more wash with RPE. RNA was then eluted, again in 30 µl water, and concentration and quality were measured again. In all cases where reprocessing was necessary, this treatment resulted in samples with ideal ratios for sequencing.

After quantification of RNA, the samples were run on a 1% agarose gel supplemented with 1% household bleach [Musgrave Retail Partners Ireland, Lucan, Dublin] to observe the integrity of the RNA preparations. Samples were deemed satisfactory if the 28S and 18S ribosomal subunit bands could be seen clearly on the gel, and that the intensity of the 28S band was more than (ideally double) that of the 18S band.

Samples were stored at -80 °C until being sent to Novogene (Beijing, China) for RNA-sequencing. 400 ng of RNA was the minimum amount of RNA required by the company. Sequencing was performed by the company using the Illumina platform.

6.2.3 RNA-sequencing data analysis

The sequencing company, Novogene, quality checked and aligned the sequencing data to the reference genome for each sample. They also quantified the expression of each gene in count and fragment per kilobase read format. For each strain they performed differential binding analysis on the sample in YEPD versus the sample under oxidative stress. They also performed Gene Ontology (GO) analysis.

The raw count data generated was used to make other comparisons, such as comparing the gene expression in the strains with the reintroduced *TLOs* versus the *ΔΔtlo* strain or the AHY940 WT strain. This was done in RStudio (www.rstudio.com) using the DESeq2 package (Love *et al.*, 2014), a package that quantifies differential gene expression. When filtering for statistical significance, cut offs of 0.05 were applied based on the adjusted *p* value outputted by the DESeq2 package. Through the DESeq2 default settings, *p* values were determined by the Wald test and then corrected for multiple testing using the Benjamini-Hochberg post-hoc test. Graphs and heatmaps were generated from the differential expression analysis in R Studio using the ggplot (Wickham, 2009) and pheatmap (https://cran.r-project.org/) packages. Lists of differentially expressed genes were analysed via GO analysis (Ashburner *et al.*, 2000) on the *Candida* Genome Database (Skrzypek *et al.*, 2017) and GSEA (Subramanian *et al.*, 2005). Venn diagrams were generated from lists using the online tool from the Gent bioinformatics group (http://bioinformatics.psb.ugent.be/webtools/Venn/).

6.3 Results

6.3.1 Sequencing quality and mapping statistics

Strains AHY940, *ΔΔtlo*, *ΔΔtlo::P_{TET}TLOα1, ΔΔtlo::P_{TET}TLOβ2* and *ΔΔtlo::P_{TET}TLOγ11* were subjected to RNA sequencing in two separate conditions, one during mid-exponential phase growth in YEPD, and the other after being grown to mid-exponential phase in YEPD then subjected to oxidative stress (exposure to 0.25 M tBOOH for 30 min). RNA samples were sequenced by Novogene on the Illumina platform. Sequencing quality scores can be found in Appendix 4. On average, 6.9 gigabases of data were obtained for each sample, an average of around 23,000,000 raw reads. No sample had an error rate higher than 0.03%. The average sample had 97% of reads with a quality score of Q20 or higher, and 93.5% with a score of Q30 or higher.

Novogene also performed some bioinformatic analysis, including the alignment of the reads to the reference genome and quantification of read counts at each gene. They also performed some differential expression analysis, comparing each strain to itself in YEPD or in oxidative stress. Further differential expression analysis was performed as part of this study. The mapping statistics of the alignments performed by Novogene can be found in Appendix 4. It was seen that for each strain, 94-97.5% of reads were mapped to the reference genome, of these 90-95% were uniquely mapped, with no strain displaying a multiple mapping rate of more than 3.99%. These statistics were satisfactory for proceeding with the analyses. Novogene also performed a principal component analysis. Figure 6.1, on the aligned files which showed the reproducibility for the majority of strains sequenced was high, with some exceptions, specifically those for the *ΔΔtlo::P_{TET}TLOγ11* strain under both YEPD and oxidative stress conditions.

6.3.2 Comparative transcriptomics of the ΔΔtlo *mutant and AHY940 in nutrient rich YEPD*

Phenotypic analysis demonstrated that there are significant differences in the morphology and growth rate of the *ΔΔtlo* mutant and its parent, AHY940. Differential gene expression analysis was performed to identify if differences in the transcriptional profiles of the two strains might help explain the molecular basis for the differences observed and therefore help us understand the function of the *TLO* family. For these

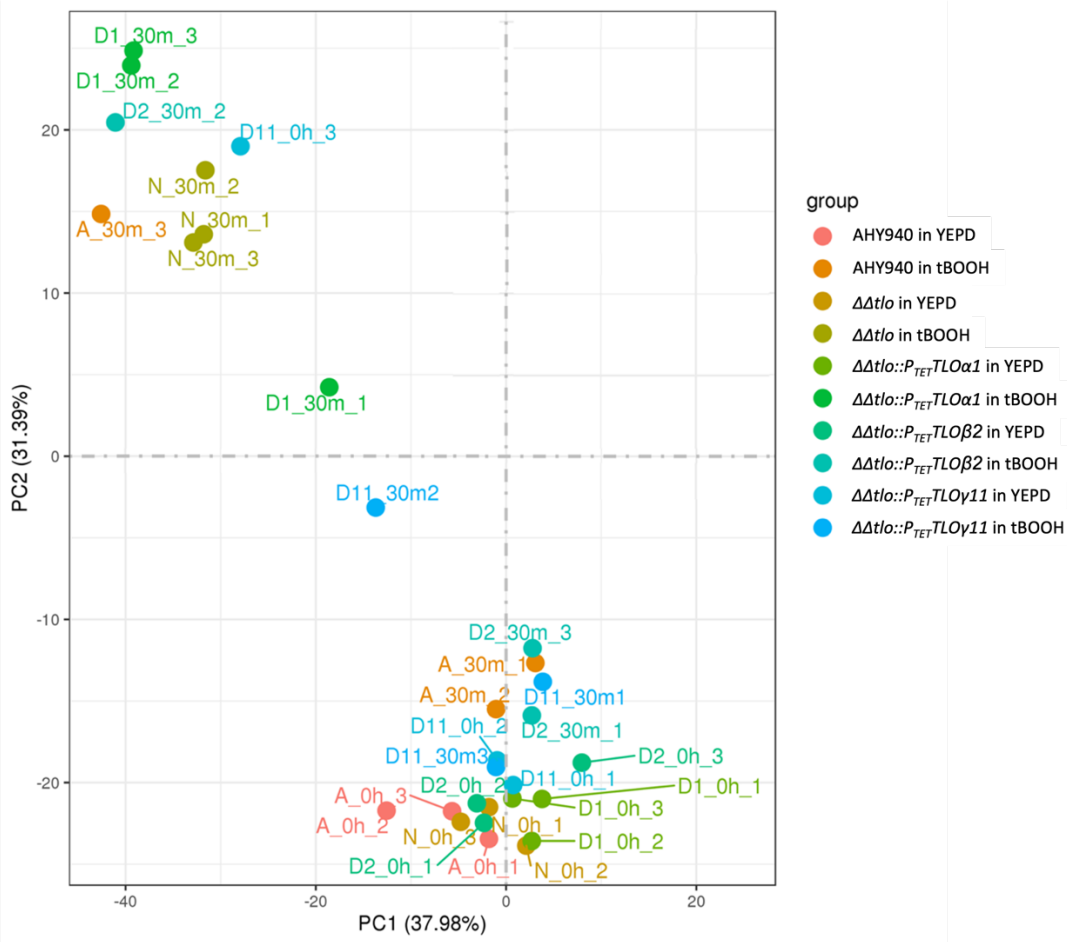


Figure 6.1 Principal Component Analysis of gene expression correlation between *TLO* mutant strains sequenced

The PCA plot is based on the gene expression correlation between samples, based on fragments per kilobase million (quantification of gene expression). Distance between samples can be used to infer how similar or distinct the samples are.

experiments, RNA was extracted from the parent and mutant strains grown in YEPD at 37 °C to mid-exponential phase.

Of 6,131 genes found to be differentially expressed, 2,946 were expressed at a higher level in the *ΔΔtlo* mutant, 239 of which were deemed significant (p =< 0.05), and of these 83 were upregulated by a $\log_2$ fold change of greater than 1.5. There were 3,185 genes expressed at a lower level in the *ΔΔtlo* mutant than the parent, with 350 of these genes being significantly downregulated, 89 of these being more than 1.5 $\log_2$ fold different. The volcano plot in Figure 6.2 highlights the change in expression in genes between these two strains, and the significance of these changes. Differential expression analysis of the *ΔΔtlo* mutant compared to the parent also confirmed that the *TLO* genes were effectively deleted. Figure 6.3 is graphical representation of the expression change of the *TLO* gene. The mean difference in $\log_2$ fold change was -6.7.

6.3.2.1 Gene Ontology analysis

Gene Ontology (GO) analysis was performed on the lists of genes that were significantly up- or downregulated in the *ΔΔtlo* mutant compared to the AHY940 parent to determine if the lists were enriched for specific GO terms. Visual representation of these results can be seen in Figure 6.4. The list of genes that were significantly upregulated in the *ΔΔtlo* mutant were found to be enriched for regulation of reproductive processes (GO:2000241) (i.e. *BEM1, CST5, CTF5, DLH1, FAB1, FAV1, HST6, RAD50, SGO1, SMC2, SPO22, STE18, SWE1, YCG1* and uncharacterised ORF C6_00540W_A), DNA recombination (GO:0006310) (i.e. *CDC45, CTF5, DLH1, MRE11, MSH2, PMS1, RAD50, RAD54, YKU80* and some uncharacterised ORFs), chromosome organisation (GO:0051276) (i.e. *CBF1, CDC13, CTF5, DLH1, ESP1, MRE11, MSH2, PMS1, RAD4, RAD50, RAD54, RFC2, RVB2, SGO1, SMC2, SPT6, YCG1, YKU80* and some uncharacterised ORFs) and ribosomal large subunit biogenesis (GO:0042273) (i.e. *DBP7, DUR4, ERB1, HIT1, JIP5, LIG1, MAK21, MSU1, NOP14, PHO89, PRP3, PRP5, RAD4, RIA1, RPR2, RRP8, SDA1, SPO22, TRK1, YCG1* along with some uncharacterised ORFs).

Genes downregulated in the *ΔΔtlo* mutant were enriched for many metabolic and biosynthetic processes, examples include GO terms such as glycolytic process (GO:0006096) (i.e. *ADH1, CDC19, ENO1, FBA1, GPM1, PFK1, PFK2, PGK1, TDH3, TPI1* and *TYE7*) and cellular alcohol biosynthesis processes (GO:0044108) (i.e. *ARC40, CBR1, COF1, CYP1, DAP1, DPP1, ENO1, ERG251, ERG3, ERV46, HEM13, HGT7, MMD1, MNT1, PDC11, PGA10, PTR22, SAR1, SCW11* and other uncharacterised ORFs).

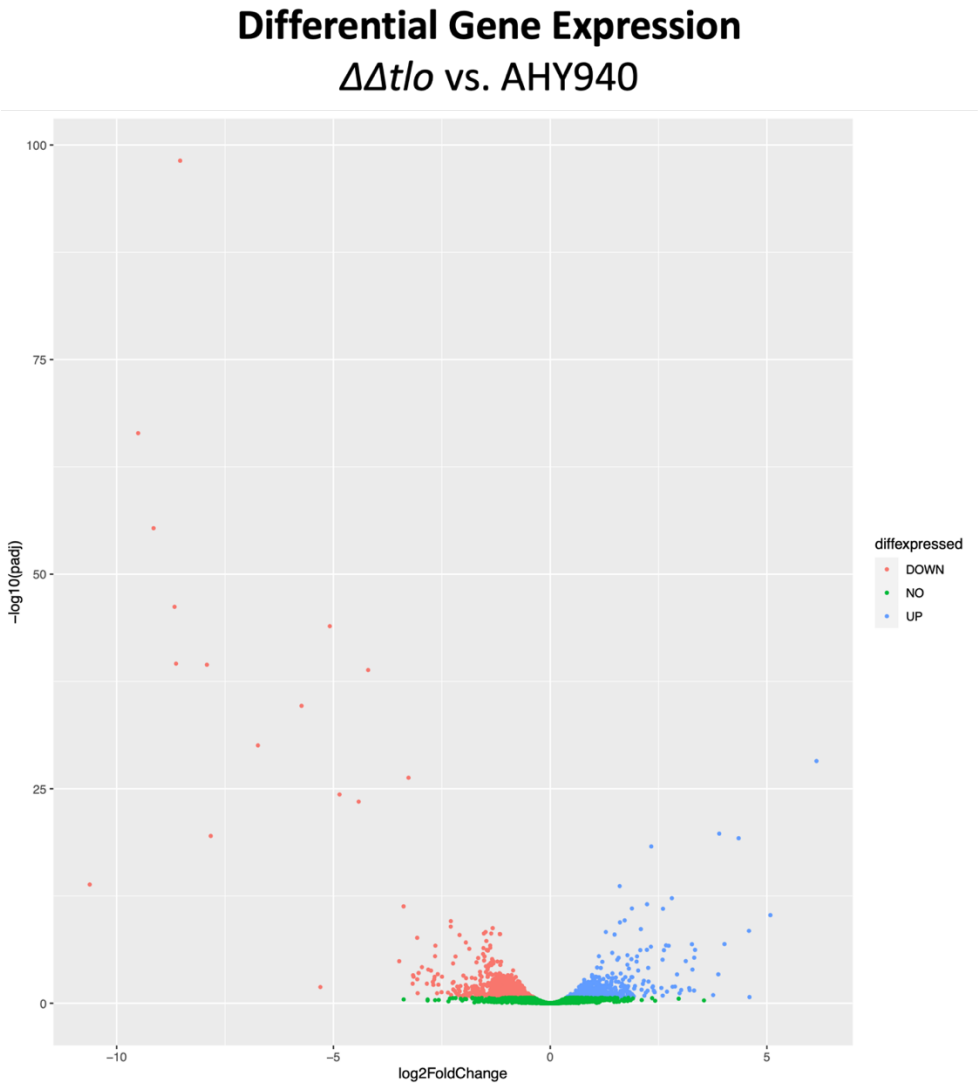


Figure 6.2 Differential gene expression in the *ΔΔtlo* mutant compared to the AHY940 WT parent grown to mid-exponential phase in YEPD at 37 °C

The $\log_2$ fold change of the expression of each gene is plotted on the x-axis, and the $-\log_{10}$ of the adjusted p value on the y-axis. Points to the right of 0 on the x-axis are expressed more highly in the *ΔΔtlo* mutant, and those to the left are expressed at a lower level in the *ΔΔtlo* mutant. Points in green ("NO") are not significant, those in blue ("UP") are significantly upregulated genes, while red ("DOWN") represents significantly downregulated genes.

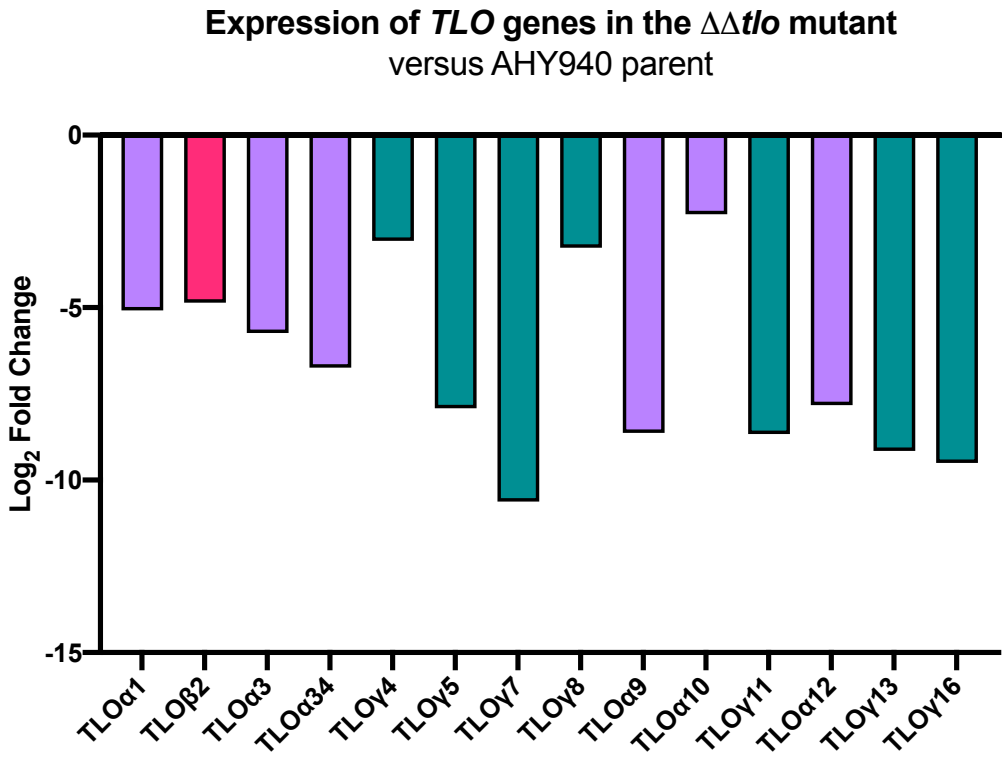


Figure 6.3 Expression of the *TLO* genes in the *ΔΔtlo* mutant compared to the AHY940 WT parent

Differential expression values (log_2 fold change) of the *TLO* genes in the *ΔΔtlo* mutant compared to the AHY940 WT parent were graphed. *TLO* clades are colour coded; α (purple), β (pink), and γ (green).

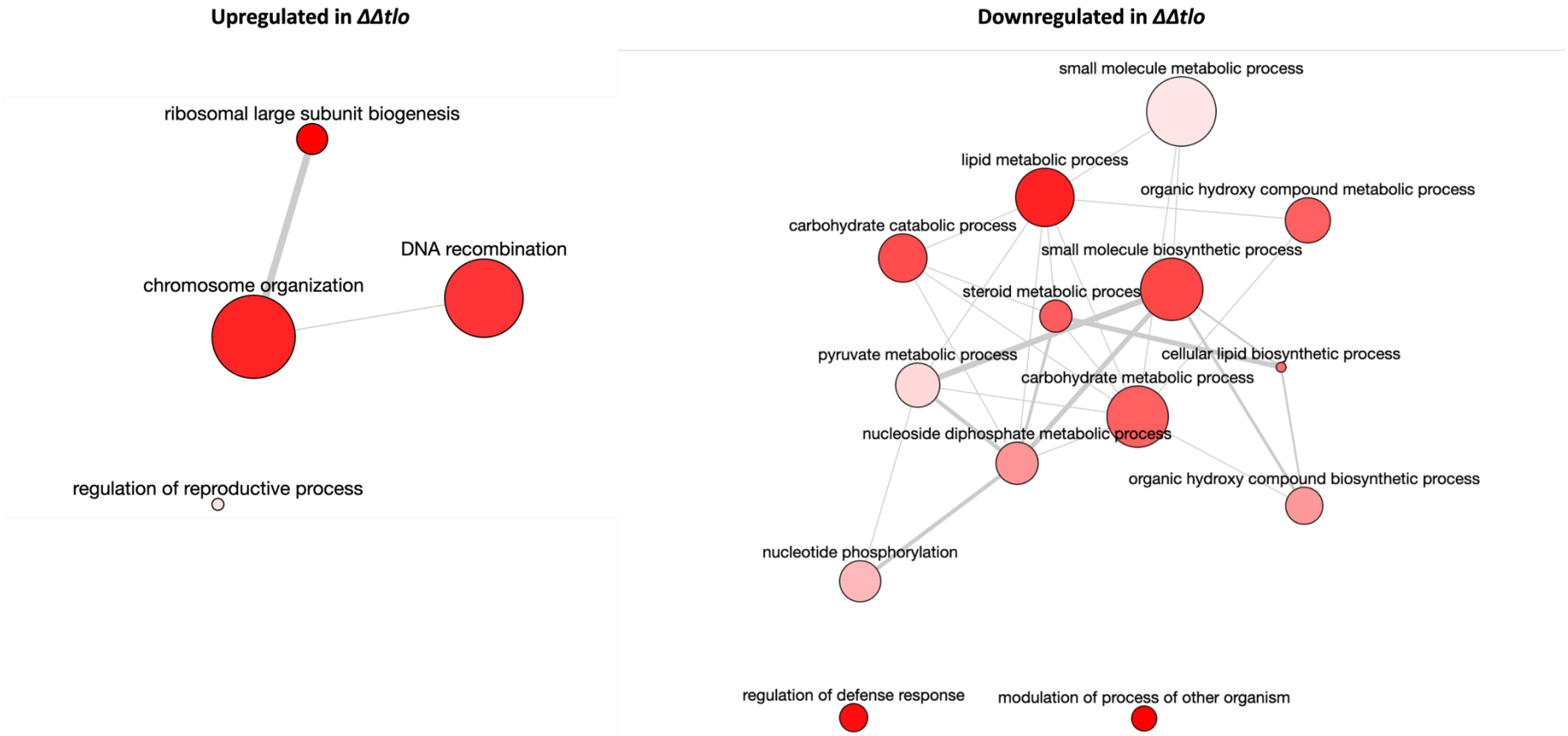


Figure 6.4 Gene Ontology analysis of genes significantly up or downregulated in the *ΔΔtlo* mutant versus the AHY940 parent

The maps above show the enriched biological process GO terms for the lists of gene either up or downregulated in the *ΔΔtlo* mutant versus AHY940 as bubbles, with the lighter colour indicating more significance. Bubbles are joined by lines, the thickness of which indicates the relatedness of the terms. Bubble size correlated to the size of the GO term in the database. Data was visualised using the online Revigo tool.

6.3.2.2 Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) was performed on the list of all differentially expressed genes, ranked by log_2 fold change, to determine if there was any enrichment of gene sets in either the parent or the null mutant. There were 232 gene sets (of 3,661) that were significantly enriched (p =< 0.05) in the parental AHY940 strain, specifically including sets of genes found on chromosome 5 (between the 1- 1,450,000 bp positions), Figure 6.5. Excluding the chromosome 5 sets, the top enriched set in the AHY940 strain was "WHITE_UP" which contains genes that are upregulated in white type cells (rather than opaque cells), Figure 6.6 (A). Other highly enriched sets included "BMDM_PHAGOCYTOSIS_UP" which contains genes upregulated upon *C. albicans* being phagocytosed by macrophages, "OS_UP" which contains genes upregulated in *C. albicans* when exposed to osmotic shock, "KETOCONAZOLE_UP" which contains genes upregulated in *C. albicans* exposed to ketoconazole (the *ERG11,* located on chromosome 5, gene is significantly upregulated in AHY940 compared to the *ΔΔtlo* mutant [+1.2 L_2FC in AHY940]), "BIOFILMBATCH_UP" which contains that are upregulated in continuous biofilm grown *C. albicans* vs. batch grown, "RAS1_YEAST_UP" which contains genes upregulated in a *ΔΔras1* mutant, "HYPHAE_FBS37(NANTEL)_DN" which contains genes from an unpublished dataset of genes downregulated in hyphae (induced by FBS and temperature increase), and "TAC1_DN" which contains genes downregulated in a *ΔΔtac1* mutant (i.e. putative Tac1 activated genes), Figure 6.6 (A). Gene sets containing genes upregulated during infection of reconstituted human epithelium at 30 min and 90 min were also enriched, "RHE30_UP" and "RHE90_UP".

In the *ΔΔtlo* mutant there were 407 significantly enriched gene sets. The top gene set enriched in the *ΔΔtlo* mutant was the "NUCLEOLUS_CEL" set, which contains genes that fall under the "Nucleolus" cellular component GO term (GO:0005730), Figure 6.6 (B). Other enriched sets included the biological process GO terms "Ribosome biogenesis and assembly" (GO:0042254). "rRNA processing" (GO:0006364), "rRNA metabolic process" (GO:0016072), and "Ribonucleoprotein complex biogenesis and assembly" (GO:0022613), as well as other gene sets containing genes encoding proteins proposed to interact with Nip7*,* Has1, Rlp7, and R5, respectively. In addition, the "UPC2_DN" set, which contains genes that are downregulated in a *ΔΔupc2* mutant (i.e. putative Upc2 activated genes), were also enriched, Figure 6.6 (B). The *UPC2* gene itself is not significantly differentially expressed in the *ΔΔtlo* and AHY940 strains, however many of the genes impacted by this transcription factor are located on chromosome 5. Significantly enriched sets in the *ΔΔtlo* mutant also highlighted enrichment of sets

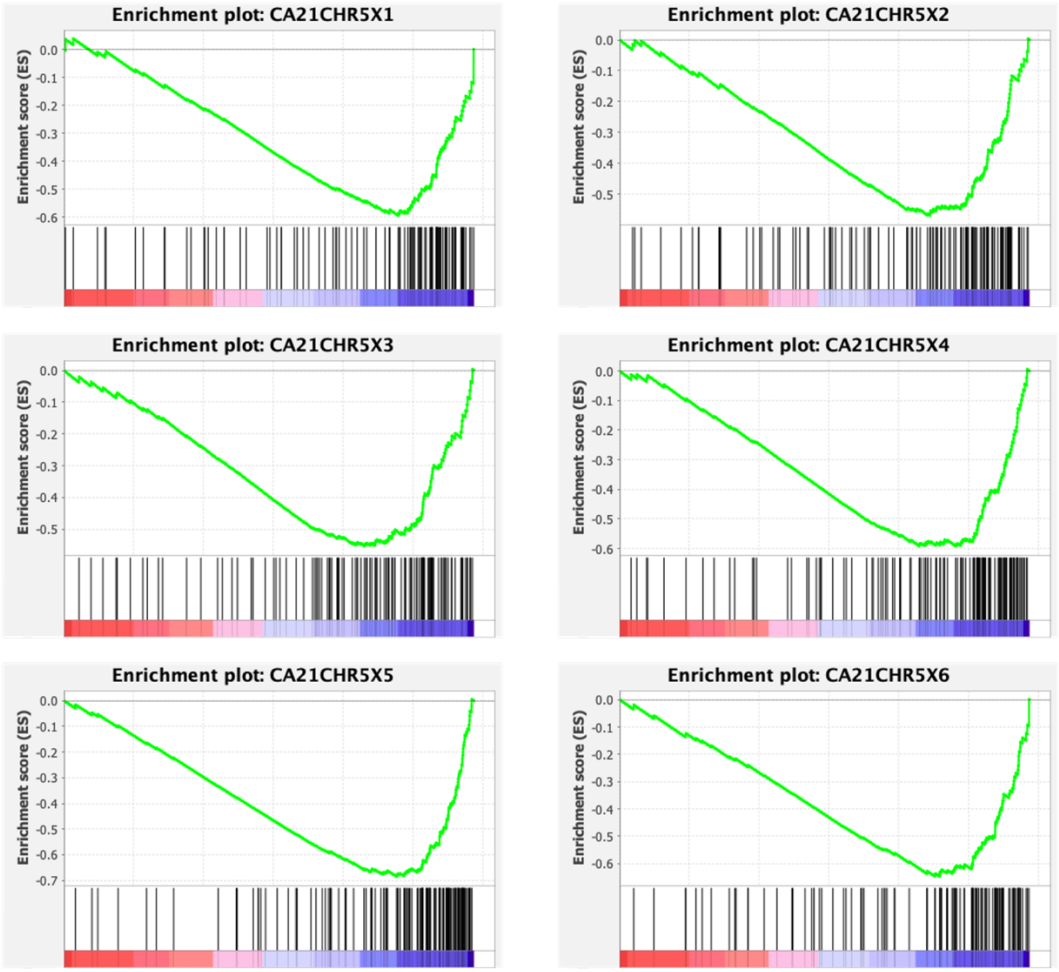


Figure 6.5 Enrichment plots of genes on chromosome 5 from comparison of the *ΔΔtlo* mutant to the AHY940 parent

Enrichment plots were generated using GSEA software, comparing the lists of genes differentially expressed between the *ΔΔtlo* mutant and the AHY940 parent (ranked by $\log_2$ fold change). Plots depict the enrichment score (top, green) for the genes across the ranked list, with the black bars indicating where each gene in the set falls in the list. Skew to the right here indicates enrichment in AHY940, as does the enrichment score being a negative number. Here, enrichment of sets of genes on chromosome 5 are displayed, where a known trisomy is present in AHY940 (resolved in *ΔΔtlo*). Each set represents 250-300 kbp of the chromosome; CA21CHR5X1 represents 1-250,000 bp, X2 = 150,000-450,000 bp, X3 = 350,000-650,000 bp, X4 = 550,000..850,000 bp, X5 = 750,000-1,050,000 bp and X6 = 950,000-1,250,000 bp.

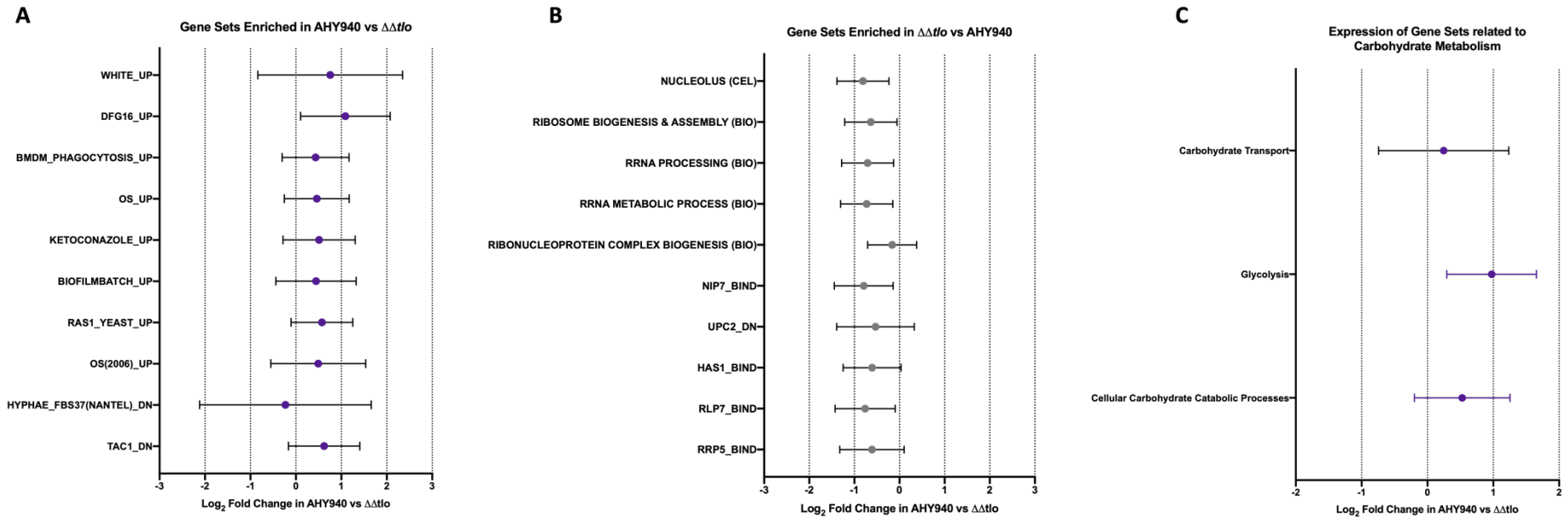


Figure 6.6 Expression of most enriched gene sets in the *ΔΔtlo* mutant or the AHY940 parent

The mean expression values (log_2 fold change in AHY940 vs. *ΔΔtlo*) and the standard deviation of individual genes in the most highly enriched gene sets identified by GSEA were plotted. (A) Expression of genes in top gene sets enriched in AHY940 versus *ΔΔtlo* (excluding chromosome 5 sets), (B) Expression of genes in top gene sets enriched in *ΔΔtlo* versus AHY940, (C) Expression of genes in sets related to carbohydrate metabolism enriched in AHY940 versus *ΔΔtlo*.

covering chromosome 7 positions 1- 1,050,000 bp, the entire length of this chromosome. Gene sets related to carbohydrate metabolism were also enriched in the AHY940 strain when compared with the mutant, with genes falling under the "Glycolysis" biological function GO term (GO:0006096) and the "Cellular carbohydrate catabolic process" biological function GO term (GO:0044275) being significantly enriched. The "Carbohydrate transport" biological function GO term (GO:0015144) was also enriched, but not significantly, Figure 6.6 (C).

6.3.3 Comparative transcriptomics of TLO *reintroduction strains and the* ΔΔtlo *mutant and in nutrient rich YEPD*

Phenotypic data described in Chapter 5 suggests that individual *TLO* genes may have specific functions. In order to investigate this further, the strains where *TLOα1, β2* or *γ11* were expressed under the *TET1* promoter in the *ΔΔtlo* mutant background were also subjected to RNA-sequencing, again being grown to mid-exponential phase in YEPD at 37 °C before RNA extraction. Differential expression analysis was performed comparing the reintroduction strains to the *ΔΔtlo* mutant. Volcano plots displaying the up- and downregulated genes and their significance in these strains can be found in Figure 6.7, and a heatmap comparing the differential expression across the three strains can be seen in Figure 6.8.

Comparing the *TLOα1* strain to the *ΔΔtlo* mutant found that 6,129 genes were differentially expressed, 3,017 genes were expressed at a higher level in the *ΔΔtlo::*P_{TET}*TLOα1* strain and 3,112 at a lower level. Of these, 1,148 genes were found to be significantly upregulated, 140 of those with a $\log_2$ fold change of more than 1.5, and 1145 were significantly downregulated, 287 with a $\log_2$ fold change of more than 1.5. In the case of the *ΔΔtlo::*P_{TET}*TLOβ2* strain in comparison to the *ΔΔtlo* mutant, 6,139 genes were differentially expressed, 3,173 were upregulated and 2,966 were downregulated. Of these, 616 were significantly upregulated, with 122 of those having a $\log_2$ fold change of more than 1.5, and 594 were significantly downregulated, with 143 of those genes having a $\log_2$ fold change of more than 1.5. For the *ΔΔtlo::*P_{TET}*TLOγ11* strain compared to the *ΔΔtlo* mutant, 6162 genes were differentially expressed, 3,274 were upregulated and 2,888 were downregulated. Of these, 44 were significantly upregulated, with 29 of those genes having a $\log_2$ fold change of more than 1.5, and 19 were significantly downregulated, with 8 of those having a $\log_2$ fold change of more than 1.5. Up- and downregulated genes in each strain were compared, Figure 6.9, and it was found that

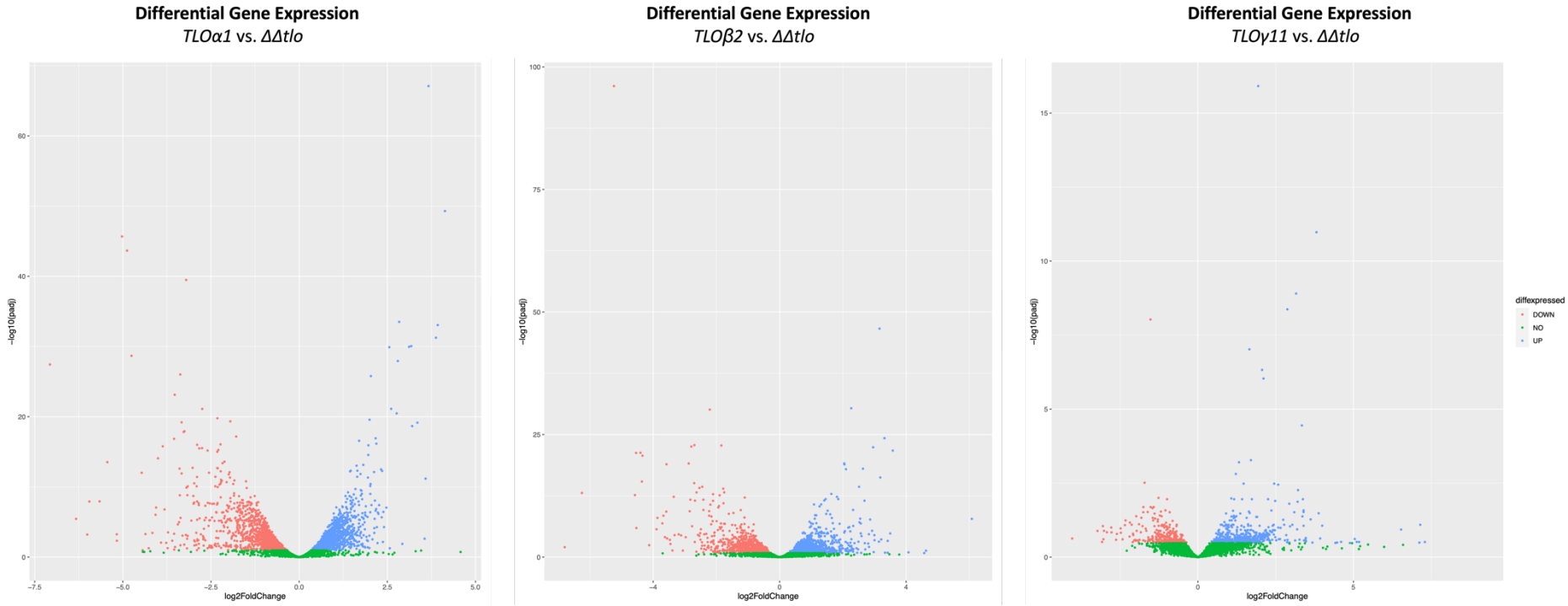


Figure 6.7 Differential gene expression in the *TLO* reintroduction strains compared to the *ΔΔtlo* mutant in mid-exponential phase in YEPD at 37 °C

The $\log_2$ fold change of the expression of each gene is plotted on the x-axis, and the $-\log_{10}$ of the adjusted p value on the y-axis. Points to the right of 0 on the x-axis are expressed more highly in the *TLO* reintroduction strain, and those to the left are expressed at a lower level in the *TLO* reintroduction strain. Points in green ("NO") are not significant, those in blue ("UP") are significantly upregulated genes, while red ("DOWN") represents significantly downregulated genes.

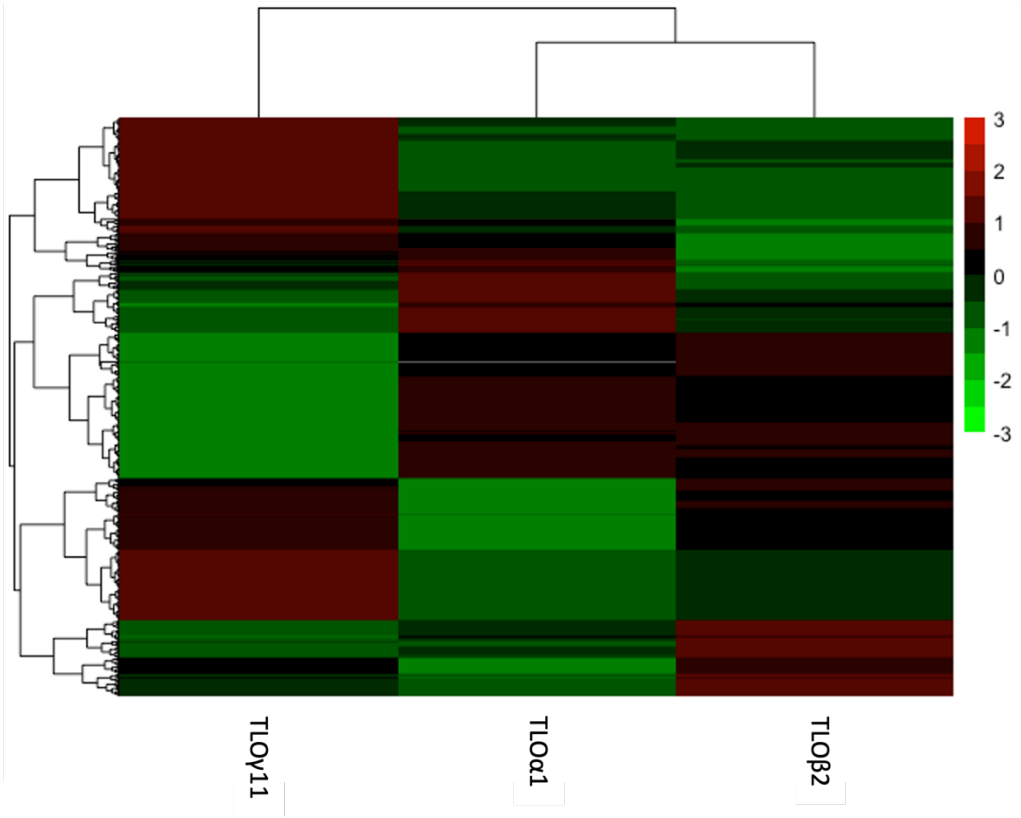
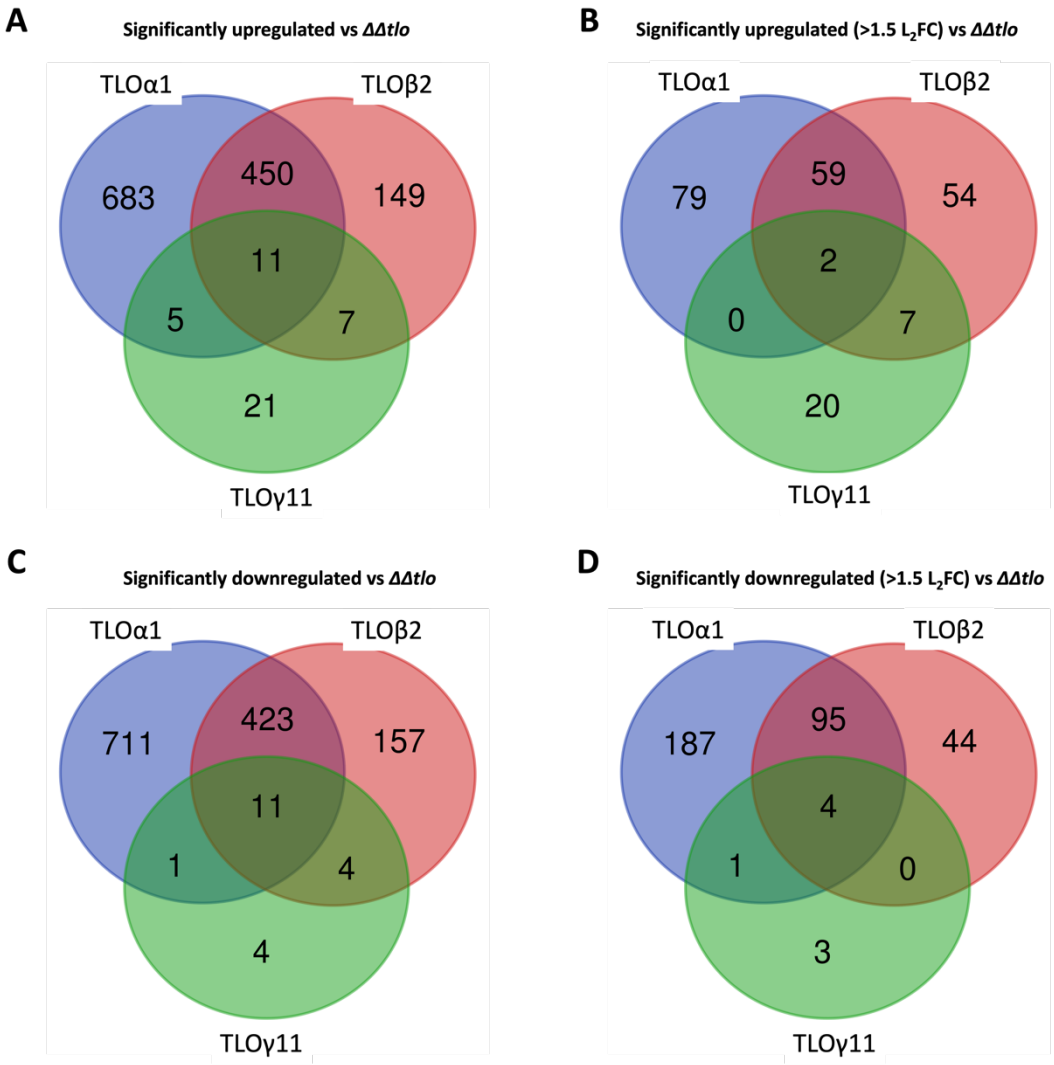


Figure 6.8 Heatmap representing differential gene expression in the *TLO* reintroduction strains compared to the *ΔΔtlo* mutant

A heatmap was generated from the $\log_2$ fold change values generated by differential gene expression analysis. Genes downregulated in a strain compared to the *ΔΔtlo* mutant are in green, while genes upregulated are red, see scale bar on right. Hierarchal clustering was performed by default on both the genes (along the y-axis) and on the strains themselves (x-axis) to represent similarity in the expression patterns.



E

	Sig. upregulated vs. *ΔΔtlo*	Sig. upregulated (>1.5 L_2FC) vs. *ΔΔtlo*	Sig. downregulated vs. *ΔΔtlo*	Sig. downregulated (>1.5 L_2FC) vs. *ΔΔtlo*
P_{TET}*TLOα1*	1,149	140	1,146	287
P_{TET}*TLOβ2*	617	122	595	143
P_{TET}*TLOγ11*	44	29	20	8

Figure 6.9 Comparison of up- and downregulated genes in *TLO* reintroduction strains versus the *ΔΔtlo* mutant

Lists of genes differentially expressed in the *TLO* reintroduction strains versus the *ΔΔtlo* mutant were compared and Venn diagrams were generated. (A) Genes significantly upregulated in *TLO* reintroduction strains, (B) Genes significantly upregulated in *TLO* reintroduction strains by $\log_2$ fold > 1.5, (C) Genes significantly downregulated in *TLO* reintroduction strains, (D) Genes significantly downregulated in *TLO* reintroduction strains by $\log_2$ fold > 1.5, (E) Summary table of total genes up- or downregulated in each *TLO* reintroduction strain.

the *ΔΔtlo::P_{TET}TLOα1* strain and the *ΔΔtlo::P_{TET}TLOβ2* strain share many of their up- and downregulated genes with each other, more than either share with the *ΔΔtlo::P_{TET}TLOγ11* strain, which was also confirmed by the hierarchical clustering in Figure 6.8.

Differential expression analysis also confirmed expression of each reintroduced *TLO* in the null background, Figure 6.10. In the *ΔΔtlo::P_{TET}TLOα1* strain, the *TLOα1* gene was expressed 4.13 $\log_2$ fold higher (p = 5.13^{-50}) than in the *ΔΔtlo* mutant. In the *ΔΔtlo::P_{TET}TLOβ2* strain, the *TLOβ2* gene was expressed 3.58 $\log_2$ fold higher (p = 1.87^{-22}) than in the *ΔΔtlo* mutant. In the case of the *ΔΔtlo::P_{TET}TLOγ11* strain the measurement of the expression of the reintroduced gene is less clean cut, likely due to ambiguous mapping of reads to other γ clade *TLO* gene transcripts that are highly similar. The *TLOγ11* gene was upregulated 1.8 $\log_2$ fold, however the p value associated with this change was not deemed significant (p = 0.4). There were other γ clade *TLO* genes that were registering as more highly expressed in the *TLOγ11* strain, such as *TLOγ4* at 0.8 $\log_2$ fold higher (p = 0.6) and *TLOγ16* at 0.2 $\log_2$ fold higher (p = 0.97) as well as *TLOα9* at 1.1 $\log_2$ fold higher (p = 0.7). The fact that no *TLO* genes are present in the *ΔΔtlo* mutant, coupled with this expression data from *ΔΔtlo::P_{TET}TLOγ11*, can be taken to mean that there is expression of *TLOγ11* in this strain and that it is possible that expression of *TLOγ11* was falsely attributed to these other transcripts.

6.3.3.1 Gene Ontology analysis

The lists of significantly up or downregulated genes in the reintroduced *TLO*-expressing strains compared to the *ΔΔtlo* mutant were subjected to GO analysis. Figure 6.11 shows graphical representation of the GO analysis (biological process) of genes significantly up- or downregulated in the *ΔΔtlo::P_{TET}TLOα1* strain compared to the *ΔΔtlo* mutant. Genes upregulated in the *ΔΔtlo::P_{TET}TLOα1* strain were found to be enriched for GO terms related to metabolism of nitrogen, carbohydrates and other metabolic processes. Specific enriched GO terms included cellular nitrogen compound metabolic process (GO:0034641) with 655 of the 1,148 genes inputted being linked to this GO term, 830 genes were linked to the primary metabolic process term (GO:004423), and 107 genes were linked to the carbohydrate derivative metabolic process term (GO:1901135). The genes that were significantly downregulated in the *ΔΔtlo::P_{TET}TLOα1* strain were enriched for GO terms related to transport (e.g. secretion, endocytosis, drug transport and carbohydrate transport) as well as autophagy. Specific GO terms associated with the downregulated genes included transport (GO:0006810) with 550 of 1,145 genes

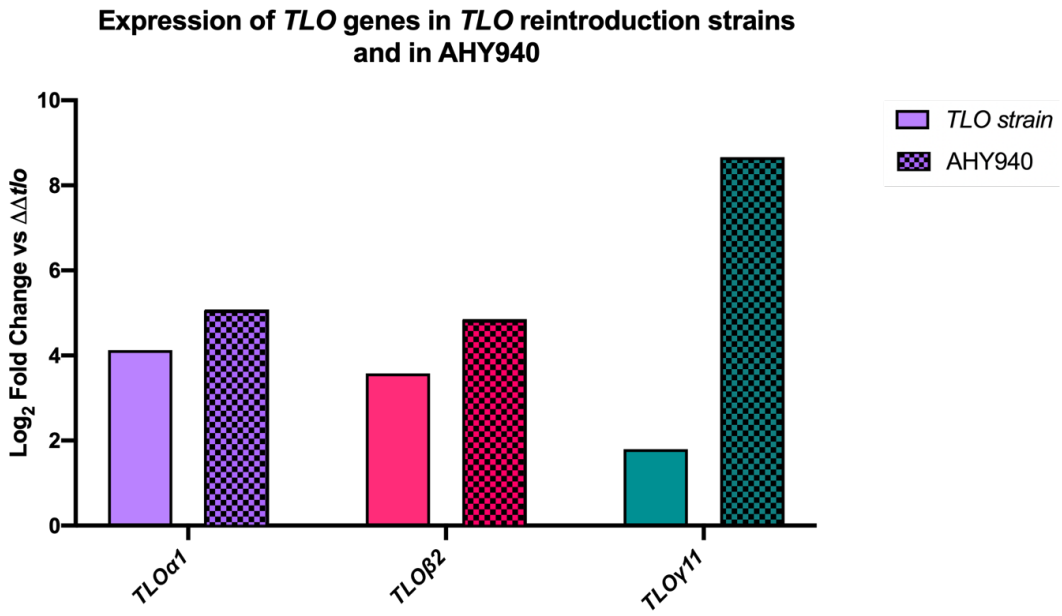


Figure 6.10 Expression of *TLO* genes in the *TLO* reintroduction strains and in AHY940

Expression change (log2 fold change) of the *TLO* genes in strains compared to the *ΔΔtlo* mutant were graphed. The expression of each *TLO* in the respective *TLO* reintroduction strain vs. the *ΔΔtlo* mutant is displayed in the plain bars; *TLOα1* in *ΔΔtlo::*P_{TET}*TLOα1* (purple), *TLOβ2* in *ΔΔtlo::*P_{TET}*TLOβ2* (pink), and *TLOγ11* in *ΔΔtlo::*P_{TET}*TLOγ11* (green). The expression of each *TLO* in AHY940 is displayed in the chequered bars; *TLOα* (purple), *TLOβ2* (pink), and *TLOγ11* (green).

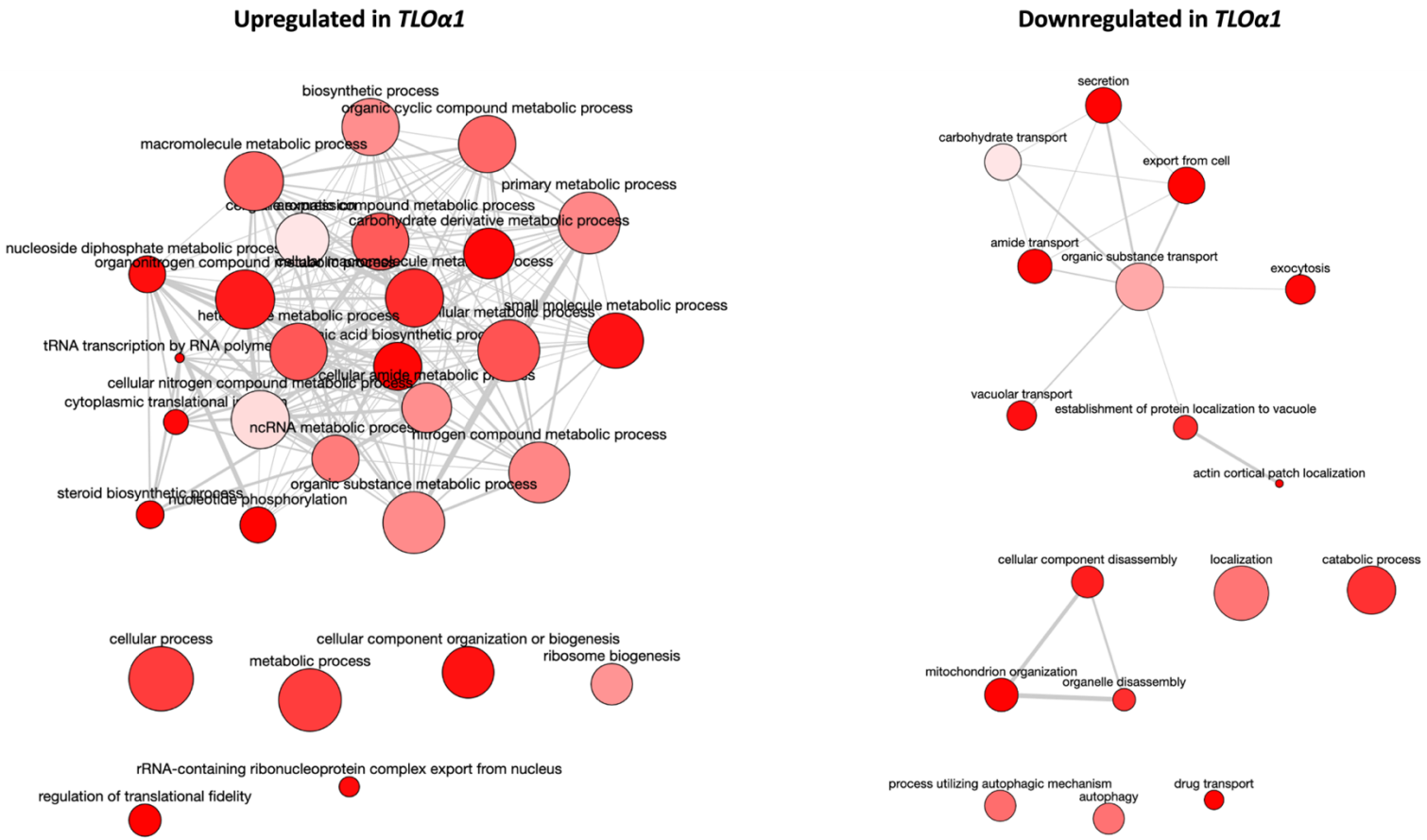


Figure 6.11 Gene Ontology analysis of genes significantly up or downregulated in the *ΔΔtlo::*P_{TET}*TLOα1* strain versus the *ΔΔtlo* mutant

The maps above show the enriched biological process GO terms for the lists of gene either up or downregulated in the *ΔΔtlo::*P_{TET}*TLOα1* strain versus the *ΔΔtlo* mutant as bubbles, with the lighter colour indicating more significance. Bubbles are joined by lines, the thickness of which indicates the relatedness of the terms. Bubble size correlated to the size of the GO term in the database. Data was visualised using the online Revigo tool.

analysed being linked to the term, 300 genes were linked to the autophagy term (GO:0006914), and 576 genes being related to the localisation term (GO:0051179).

The genes significantly up- or downregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain compared to the *ΔΔtlo* mutant were also subjected to GO analysis. Visual representation of the results can be found in Figure 6.12. It was found that the up- and downregulated genes in the *ΔΔtlo::P_{TET}TLOβ2* strain were similar to what was seen in the *TLOα1* gene lists above. Genes that were upregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain compared to the *ΔΔtlo* mutant were enriched for GO terms related to metabolism and gene expression, while those that were downregulated were related to transport, secretion, and endocytosis, as well as autophagy. Specific GO terms that were enriched in the upregulated genes included primary metabolic process (GO:004423) (416 genes out of 616 analysed), 398 genes were linked to the nitrogen compound metabolic process term (GO:0034641), and 158 genes were linked to translation (GO:0006412). Terms enriched in the downregulated genes included autophagy (GO:0006914) with 167 of 594 being related to the term and 324 being related to localisation (GO:0051179).

Gene Ontology analysis was also performed on the (comparatively small number of) genes up or downregulated in the *ΔΔtlo::P_{TET}TLOγ11* strain compared to the *ΔΔtlo* mutant. This found that the genes upregulated in the *ΔΔtlo::P_{TET}TLOγ11* strain only showed an enrichment for GO terms related to the lysine metabolic process (GO:0006553) and lysine biosynthetic process terms (GO:0009085) with the genes falling into both of these terms being *LYS12, LYS4* and *LYS9* (3 of 44 genes). There was no enrichment found when the list of downregulated genes was analysed (29 genes).

Gene Ontology analysis was also performed on the lists of genes generated by comparing the three *TLO*-expressing strains described above, Figure 6.9. Of all genes that were significantly upregulated in the strains, only 11 genes were shared between all three strains, *CSH1, CTP1, GCY1, PGA23, PGA57, PHR2, RBT7, SEO1* and three uncharacterised ORFs, Figure 6.9 (A). Of these only two genes, *PGA57* and *RBT7* were upregulated by a $\log_2$ fold change of 1.5 or more in all three strains, Figure 6.9 (B). Genes that were significantly upregulated in all three *TLO* strains were subjected to GO analysis. No enrichment was found for biological process, but it was seen that the molecular function GO terms for secondary active transmembrane transporter activity (GO:0015291) containing genes *CTP1* and *SEO1*, and D-threo-aldose 1-dehydrogenase activity (GO:0047834) containing genes *CSH1* and *GCY1*, were enriched as well as

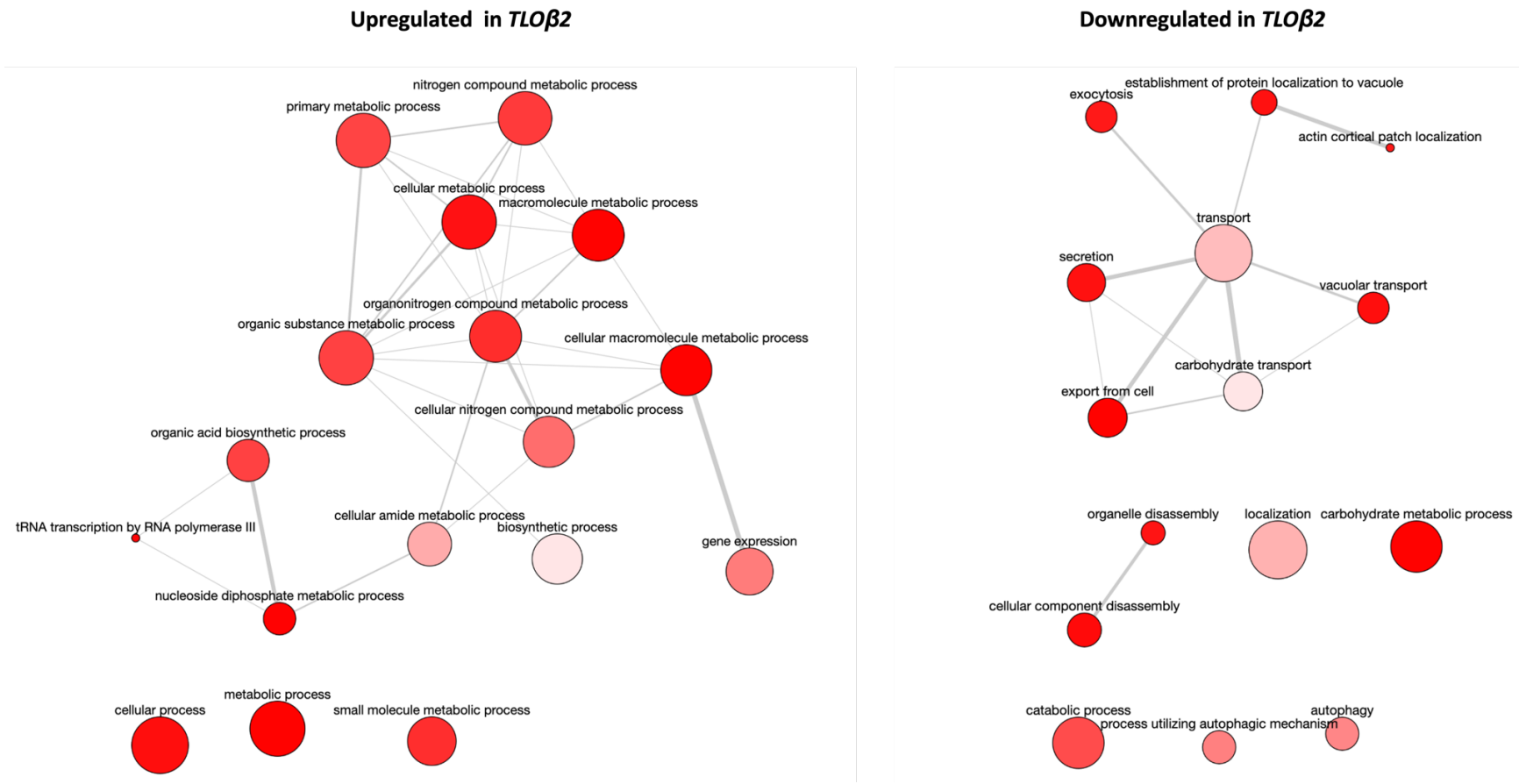


Figure 6.12 Gene Ontology analysis of genes significantly up or downregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain versus the *ΔΔtlo* mutant

The maps above show the enriched biological process GO terms for the lists of gene either up or downregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain versus the *ΔΔtlo* mutant as bubbles, with the lighter colour indicating more significance. Bubbles are joined by lines, the thickness of which indicates the relatedness of the terms. Bubble size correlated to the size of the GO term in the database. Data was visualised using the online Revigo tool.

cellular component GO terms extracellular region (GO:0005576) containing genes *PHR2*, *GCY1* and *RBT7*, cell surface (GO:0009986) containing *PHR2*, *CSH1*, *PGA57* and *PGA23*, and external encapsulating structure (GO:0030312) containing genes *PHR2*, *CSH1* and *GCY1*, were enriched in this set of genes.

In the case of the downregulated genes, there were 11 genes that were significantly downregulated in all three *TLO* strains, *LYS142, POL93, RGA2, STE11, YHB5, ZCF2* and five uncharacterised ORFs, Figure 6.9 (C). Of these, four genes, *POL93* and three uncharacterised ORFs, were downregulated by a 1.5 $\log_2$ fold change or more in all three strains, Figure 6.9 (D). The list of significantly downregulated genes that were common to all three strains was subjected to GO analysis, and while there was no enrichment for biological process or cellular component GO terms, the molecular function GO term transition metal ion binding (GO:0046914) was enriched, with genes *LYS142, RGA2, YHB5* and *ZCF2* contributing to the enrichment.

6.3.3.2 Gene Set Enrichment Analysis

Gene Set Enrichment analysis was also carried out on the lists of differentially expressed genes, ranked by $\log_2$ fold change, for each *TLO* reintroduction strain compared to the *ΔΔtlo* mutant in order to determine if any gene sets were enriched in any strains. In the case of the *ΔΔtlo::*P_{TET}*TLOα1* strain, 1,007 (of 3,660) gene sets were enriched in this strain compared to the *ΔΔtlo* mutant, while 114 sets were enriched in the *ΔΔtlo* mutant. The top enriched set in the *ΔΔtlo::*P_{TET}*TLOα1* strain was the "UPC2_DN" set, containing genes that are downregulated in a *ΔΔupc2* mutant strain of *C. albicans* (the *UPC2* gene is significantly increased in *ΔΔtlo::*P_{TET}*TLOα1* [+0.9 L_2FC]), Figure 6.13. Other enriched sets were "KETOCONAZOLE_DN", which contains genes downregulated in cells exposed to ketoconazole, Figure 6.14, and sets containing genes downregulated upon heat shock or osmotic shock, "HS_DN" and "OS(2006)_DN" respectively. The *ΔΔtlo::*P_{TET}*TLOα1* strain was also enriched for genes in the glycolysis set, "GLYCOLYSIS_BIO", Figure 6.15. Top sets enriched comparatively in the *ΔΔtlo* mutant were those containing genes upregulated in the opaque cell type, "OPAQUE_UP", Figure 6.16, as well as "KETOCONAZOLE_UP", containing genes upregulated upon exposure to ketoconazole, Figure 6.14, (the *ERG11* gene is significantly increased in *ΔΔtlo::*P_{TET}*TLOα1* [+0.9 L_2FC]) and "LIPASE ACTIVITY". There was enrichment for genes that are upregulated in a *ΔΔupc2* mutant, "UPC2_UP", as well as enrichment for sets of genes that are upregulated upon phagocytosis by macrophages, "BMDM PHAGOCYTOSIS_UP", and genes upregulated during infection of reconstituted human epithelial cells, "RHE90_UP".

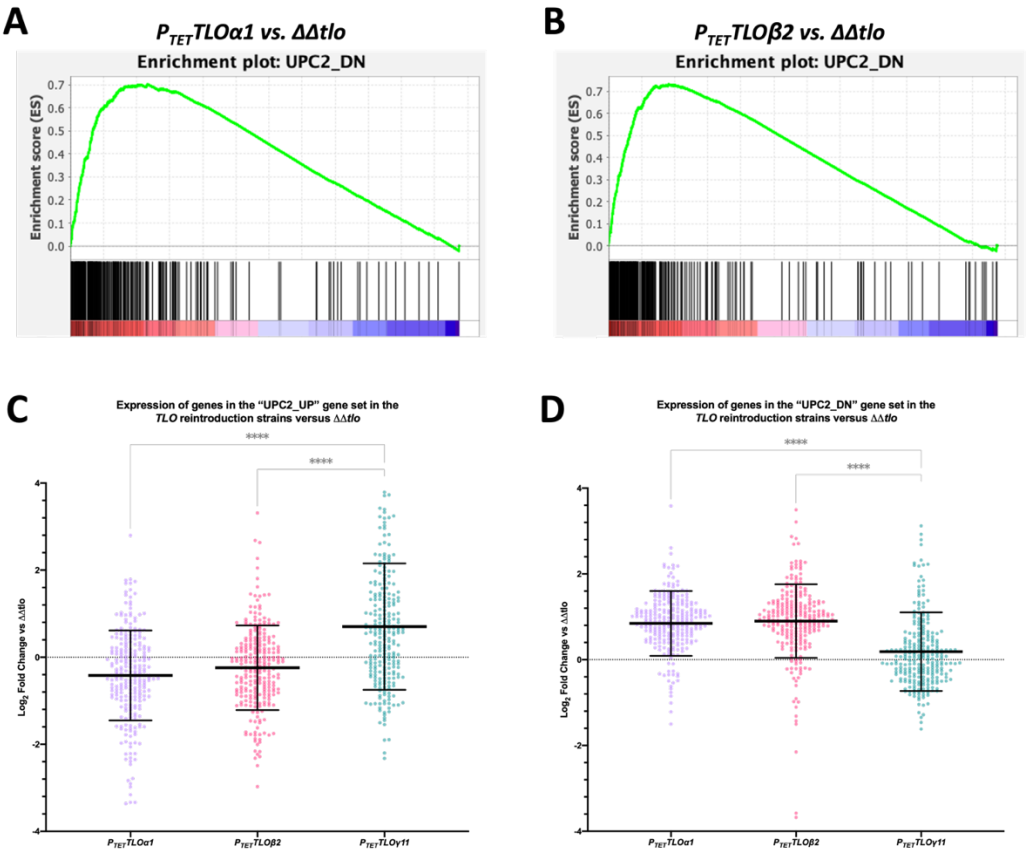


Figure 6.13 Expression of "UPC2_UP" and "UPC2_DN" gene sets in *TLO* reintroduction strains versus the *ΔΔtlo* mutant

Gene set enrichment analysis indicated that expression of genes in the "UPC2_DN" gene set (genes downregulated in a *ΔΔupc2* mutant) were enriched in both the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains compared to the *ΔΔtlo* mutant. (A) Enrichment plot for "UPC2_DN" in *ΔΔtlo::P_{TET}TLOα1* vs. *ΔΔtlo,* skew left indicates enrichment in *ΔΔtlo::P_{TET}TLOα1,* (B) Enrichment plot for "UPC2_DN" in *ΔΔtlo::P_{TET}TLOβ2* vs. *ΔΔtlo,* skew left indicates enrichment in *ΔΔtlo::P_{TET}TLOβ2*. Expression values (log_2 fold change vs. *ΔΔtlo*) were plotted for each gene in the sets "UPC2_UP" and "UPC2_DN" for each of the strains; *ΔΔtlo::P_{TET}TLOα1* (purple), *ΔΔtlo::P_{TET}TLOβ2* (pink), and *ΔΔtlo::P_{TET}TLOγ11* (green), asterisks indicate significance. (C) Expression of genes in the "UPC2_UP" gene set (genes upregulated in a *ΔΔupc2* mutant), (D) Expression of genes in the "UPC2_DN" gene set.

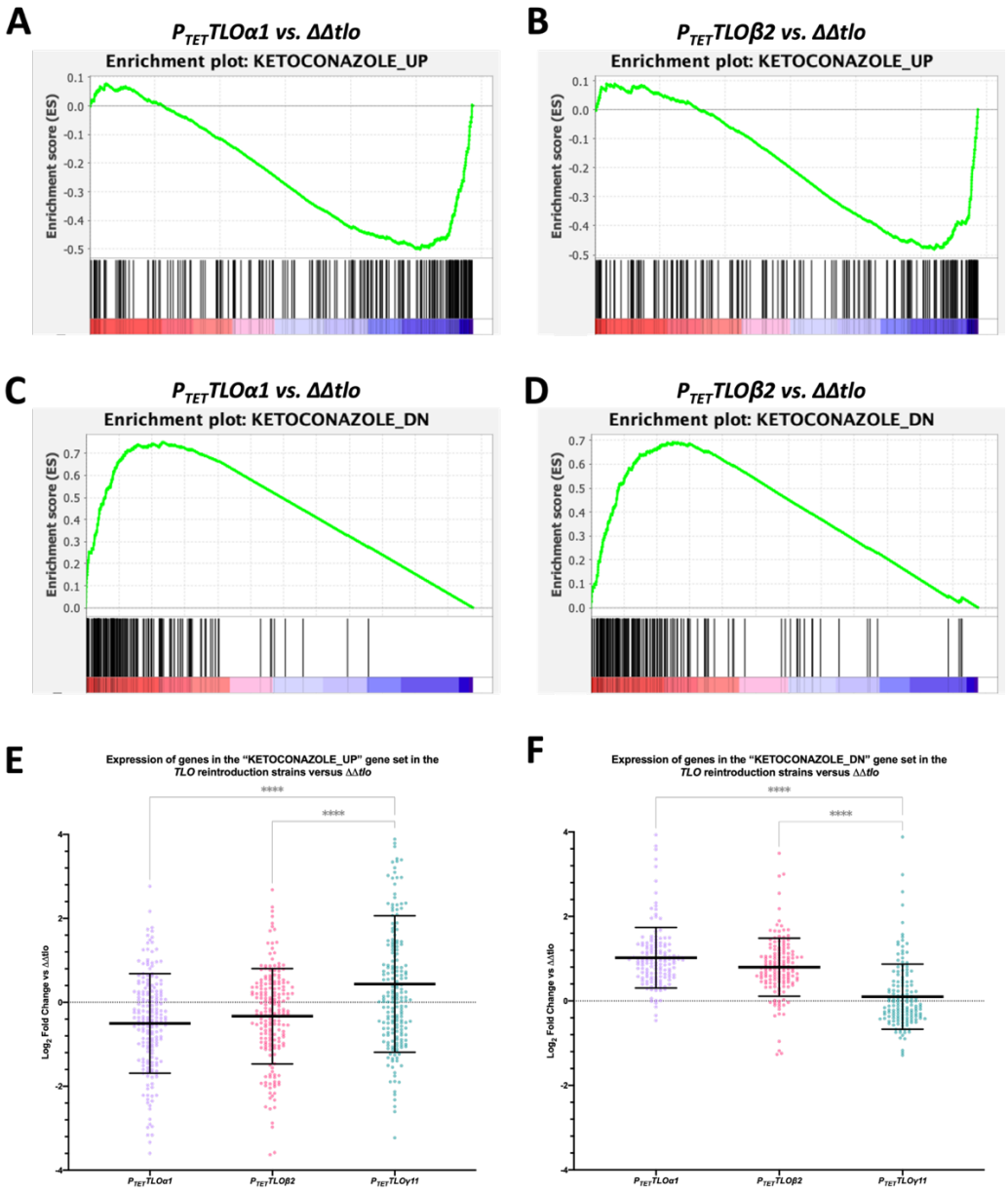


Figure 6.14 Expression of "KETOCONAZOLE_UP" and "KETOCONAZOLE_DN" gene sets in *TLO* reintroduction strains versus the *ΔΔtlo* mutant

Gene set enrichment analysis indicated the "KETOCONAZOLE_DN" gene set (genes downregulated upon exposure to ketoconazole) was enriched in both the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains vs. *ΔΔtlo*, and that the "KETOCONAZOLE_UP" (genes upregulated upon exposure to ketoconazole) set was conversely enriched in the *ΔΔtlo* mutant. (A) Enrichment plot for "KETOCONAZOLE_UP" in *ΔΔtlo::PTETTLOα1* vs. *ΔΔtlo*, skew right indicates enrichment in *ΔΔtlo,* (B) Enrichment plot for "KETOCONAZOLE_UP" in *ΔΔtlo::P_{TET}TLOβ2* vs. *ΔΔtlo,* skew right indicates enrichment in *ΔΔtlo,* (C) Enrichment plot for "KETOCONAZOLE_DN" in *ΔΔtlo::P_{TET}TLOα1* vs. *ΔΔtlo,* skew left indicates enrichment in *ΔΔtlo::P_{TET}TLOα1* (D) Enrichment plot for "KETOCONAZOLE_DN" in *ΔΔtlo::P_{TET}TLOβ2* vs. *ΔΔtlo,* skew left indicates enrichment in *ΔΔtlo::P_{TET}TLOβ2*. Expression values (log_2 fold change vs. *ΔΔtlo*) were plotted for each gene in the sets for each of the strains; *ΔΔtlo::P_{TET}TLOα1* (purple), *ΔΔtlo::P_{TET}TLOβ2* (pink), and *ΔΔtlo::P_{TET}TLOγ11* (green), asterisks indicate significance. (E) Expression of genes in the "KETOCONAZOLE_UP" set, (F) Expression of genes in the "KETOCONAZOLE_DN" gene set.

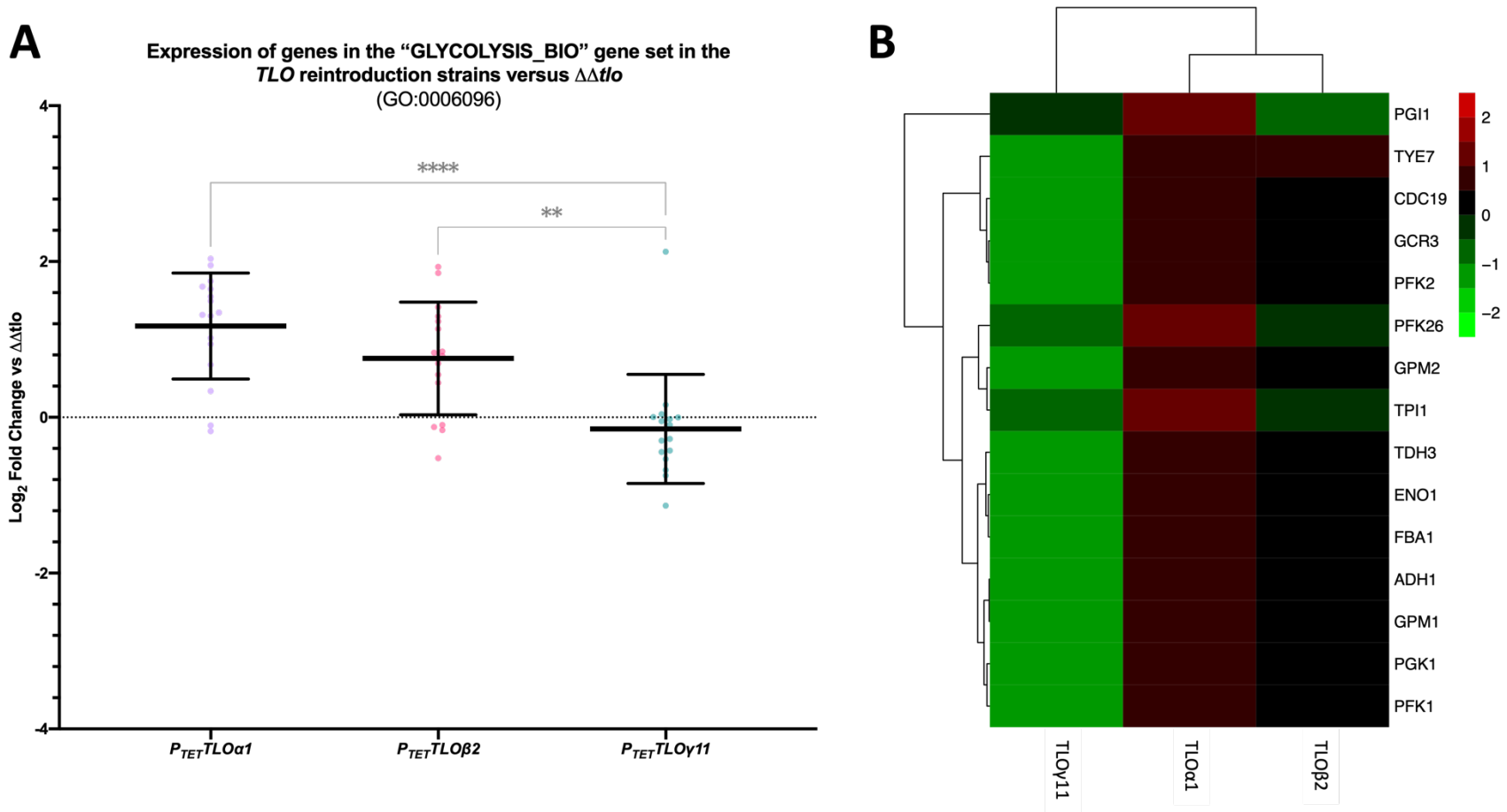


Figure 6.15 Expression of the "GLYCOLYSIS" gene set in *TLO* reintroduction strains versus the *ΔΔtlo* mutant

Expression values (log_2 fold change vs. *ΔΔtlo*) were plotted for each gene in the "GLYCOLYSIS_BIO" (GO:0006069) set (genes involved in the glycolysis pathway) for each of the strains; *ΔΔtlo::P_{TET}TLOα1* (purple), *ΔΔtlo::P_{TET}TLOβ2* (pink), and *ΔΔtlo::P_{TET}TLOγ11* (green), asterisks indicate significance. (A) Expression of genes in the "GLYCOLYSIS_BIO" set, (B) Heatmap representing the log_2 fold change of each gene in the set for each *TLO* reintroduction strain compared to the *ΔΔtlo* mutant.

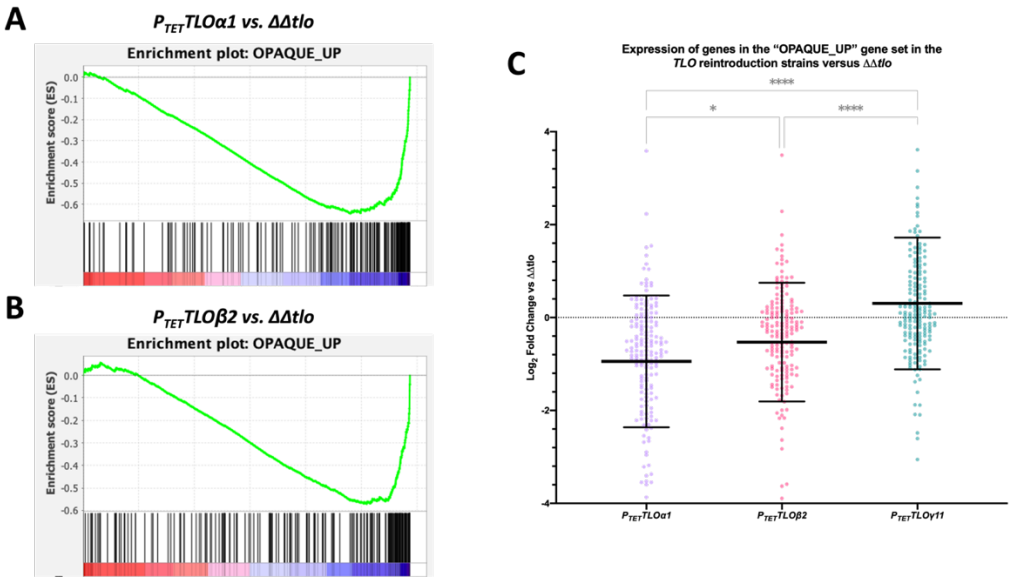


Figure 6.16 Expression of the "OPAQUE_UP" gene sets in *TLO* reintroduction strains versus the *ΔΔtlo* mutant

Gene set enrichment analysis indicated that expression of genes in the "OPAQUE_UP" gene set (genes upregulated in opaque cells) were enriched in the *ΔΔtlo* mutant compared to both the *ΔΔtlo::*P_{TET}*TLOα1* and *ΔΔtlo::*P_{TET}*TLOβ2* strains. (A) Enrichment plot for "OPAQUE_UP" in *ΔΔtlo::*P_{TET}*TLOα1* vs. *ΔΔtlo,* skew right indicates enrichment in *ΔΔtlo,* (B) Enrichment plot for "OPAQUE_UP" in *ΔΔtlo::*P_{TET}*TLOβ2* vs. *ΔΔtlo,* skew right indicates enrichment in *ΔΔtlo,* (C) Expression values (log_2 fold change vs. *ΔΔtlo*) were plotted for each gene in the "OPAQUE_UP" set for each of the strains; *ΔΔtlo::*P_{TET}*TLOα1* (purple), *ΔΔtlo::*P_{TET}*TLOβ2* (pink), and *ΔΔtlo::*P_{TET}*TLOγ11* (green), asterisks indicate significance.

There was also enrichment of "MED31_UP" which contains genes that are upregulated in a *ΔΔmed31* mutant.

Comparison of the *ΔΔtlo::P_{TET}TLOβ2* strain to the *ΔΔtlo* mutant highlighted that 689 gene sets (of 3,660) were significantly enriched in the *ΔΔtlo::P_{TET}TLOβ2* strain, and 176 sets were comparatively enriched in the *ΔΔtlo* mutant. The top set enriched in the *ΔΔtlo::P_{TET}TLOβ2* strain was the "UPC2_DN" set, Figure 6.13, (expression of the *UPC2* gene is increased in *ΔΔtlo::P_{TET}TLOβ2* [+0.3 L_2FC] but did not have a significant *p* value) Sets "HS_DN", "OS_DN", "OS(2006)_DN" and "KETOCONAZOLE_DN", Figure 6.14, were also enriched in the *ΔΔtlo::P_{TET}TLOβ2* strain. There was also enrichment for the glycolysis gene set, "GLYCOLYSIS_BIO". Genes *PFK26* and *TPI1* were highlighted as these were the only genes in this set which appeared downregulated compared to the *ΔΔtlo* mutant in *ΔΔtlo::P_{TET}TLOβ2*, but were upregulated in the *ΔΔtlo::P_{TET}TLOα1* strain. The heatmap in Figure 6.17 does not exclude non-significant changes in gene expression, therefore the different levels of expression of *PGI1* are highlighted, although the change in expression compared to the *ΔΔtlo* mutant is not significant in either *ΔΔtlo::P_{TET}TLOα1* or *ΔΔtlo::P_{TET}TLOβ2.* The expression of *PEK26* is not significantly different in *ΔΔtlo::P_{TET}TLOα1* compared to the *ΔΔtlo* mutant, but this gene is downregulated in *ΔΔtlo::P_{TET}TLOβ2* (-0.5 L_2FC with a *p* value slightly exceeding the threshold of 0.05 [p = 0.0529]). The *TPI1* gene is significantly upregulated in the *ΔΔtlo::P_{TET}TLOα1* strain (+0.9 L_2FC vs. *ΔΔtlo*), however, this gene is not significantly differentially regulated in *ΔΔtlo::P_{TET}TLOβ2* compared the *ΔΔtlo* mutant. The *ΔΔtlo* mutant strain was enriched for gene sets containing genes related to the opaque cell type, "OPAQUE_UP" and "OPAQUE_2_UP", Figure 6.16, as well as "KETOCONAZOLE_UP", Figure 6.14, (the *ERG11* gene is increased in *ΔΔtlo::P_{TET}TLOβ2* [+0.4 L_2FC] but did not have a significant *p* value) and "LIPASE ACTIVITY". There was enrichment of sets "BMDM PHAGOCYTOSIS_UP", "RHE90_UP" and "RHE30_UP". The "MED31_UP" gene set was also enriched. There was a variety of gene sets related to the hyphal morphology found to be enriched in either the *ΔΔtlo::P_{TET}TLOβ2* strain or the *ΔΔtlo* mutant in this comparison. There were both gene sets containing genes upregulated in hyphae, and gene sets containing genes downregulated in hyphae found enriched in the *ΔΔtlo::P_{TET}TLOβ2* strain, and similar enrichment of both up- and downregulated gene sets was also found in the *ΔΔtlo* mutant, without any clear consensus found. This may be due to the mixed cell morphology of the *ΔΔtlo::P_{TET}TLOβ2* strain in which only a subset of cells are hyphal under non-inducing conditions.

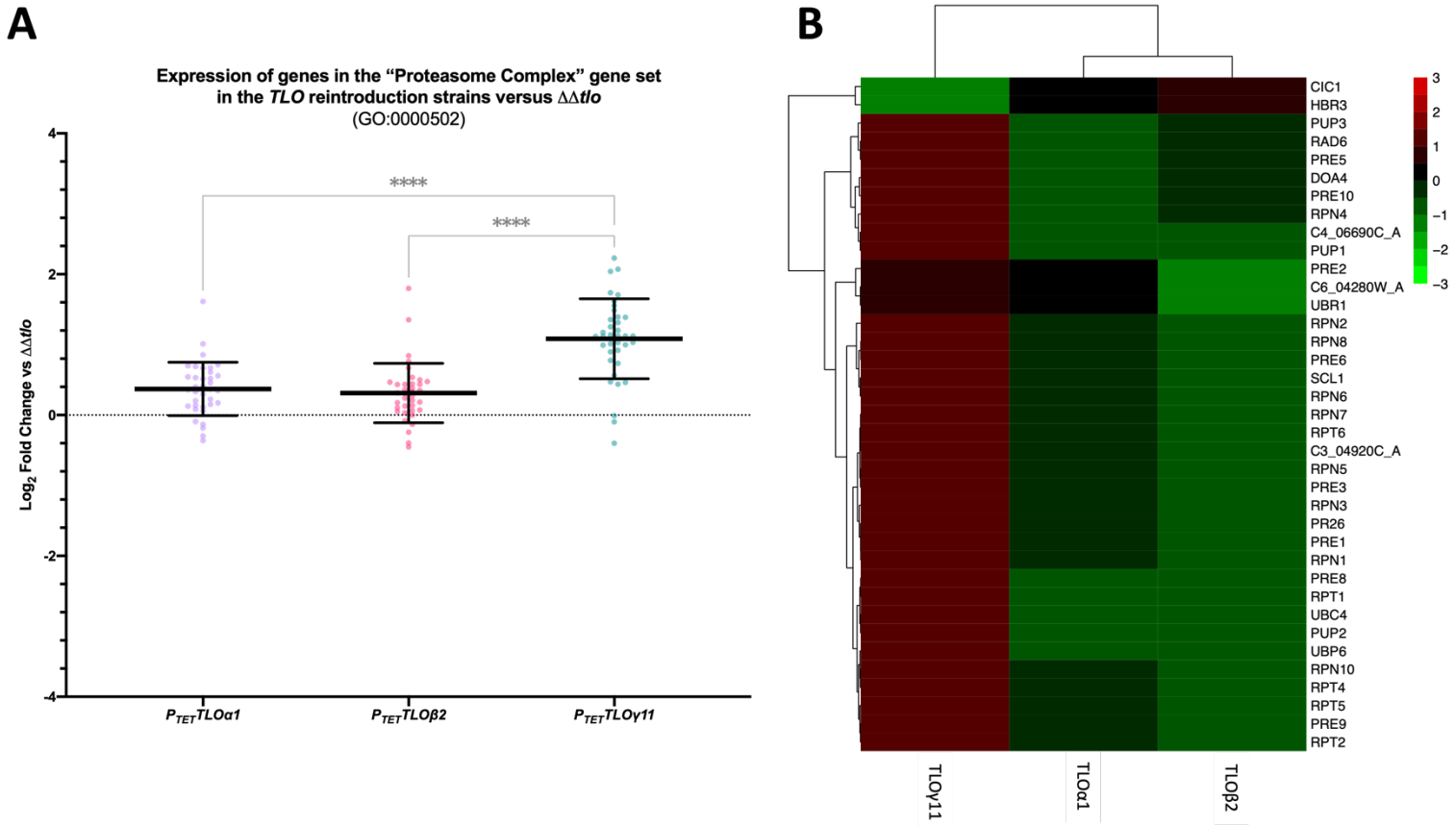


Figure 6.17 Expression of genes in the Proteasome Complex GO term (GO:0000502) in *TLO* reintroduction strains versus the *ΔΔtlo* mutant

Expression values (log_2 fold change vs. *ΔΔtlo*) were plotted for each gene in the "PROTEASOME COMPLEX" (GO:0006069) set for each of the strains; *ΔΔtlo::*P_{TET}*TLOα1* (purple), *ΔΔtlo::*P_{TET}*TLOβ2* (pink), and *ΔΔtlo::*P_{TET}*TLOγ11* (green), asterisks indicate significance. (A) Expression of genes in the "PROTEASOME COMPLEX" set. (B) Heatmap representing the log_2 fold change of each gene in the set for each *TLO* reintroduction strain compared to the *ΔΔtlo* mutant.

Comparing the lists of significantly enriched gene sets in each of *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* versus the *ΔΔtlo* mutant, it was found that of the 1,007 gene sets enriched in *ΔΔtlo::P_{TET}TLOα1* and the 689 enriched in *ΔΔtlo::P_{TET}TLOβ2,* 622 of these were enriched in both strains compared to the *ΔΔtlo* mutant. There were 385 gene sets uniquely enriched in the *ΔΔtlo::P_{TET}TLOα1* strain, and 67 uniquely enriched in the *ΔΔtlo::P_{TET}TLOβ2* strain. In the *ΔΔtlo::P_{TET}TLOα1* strain, many gene sets that were identified in *S. cerevisiae* ChIP-chip microarrays were enriched, however many of these proteins have homologues in *C. albicans*. These gene sets included genes encoding proteins bound by Erg3, Erg5 and Erg28 proteins. There was also enrichment in the *ΔΔtlo::P_{TET}TLOα1* strain of genes controlled by transcription factors *TYE7* and *GAL4.* The increased expression of genes influenced by these transcription factors can be seen in Figure 6.15 (A&B), as *TYE7* and *GAL4* are key transcriptional regulators of in the carbohydrate metabolism and glycolysis pathways. Expression of *TYE7* is significantly increased in both the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains (+2 and +1.9 L_2FC vs. *ΔΔtlo* respectively), however the expression of *GAL4* was not significantly altered in either strain compared to the *ΔΔtlo* mutant.

Compared to *ΔΔtlo::P_{TET}TLOβ2,* the gene sets "HYPHAE_FBS_37_DN" and "HYPHAE FBS37 (NANTEL)_DN" were uniquely enriched in the *ΔΔtlo::P_{TET}TLOα1* strain versus the *ΔΔtlo* mutant. In the *ΔΔtlo::P_{TET}TLOβ2* strain, the gene set bound by Ume6 was uniquely enriched (although expression of *UME6* itself was downregulated [-2 L_2FC]compared to the *ΔΔtlo* mutant), as well as gene sets bound by Ctf8, Dcc1 and Rad57 (all identified via ChIP-chip in *S. cerevisiae*).

The *ΔΔtlo::P_{TET}TLOγ11* strain was also subjected to GSEA to compare this strain to the *ΔΔtlo* mutant. There were 219 gene sets (of 3,660) found to be enriched in the *ΔΔtlo::P_{TET}TLOγ11* strain, and conversely 110 gene sets were enriched in the *ΔΔtlo* mutant. Some of the top sets enriched in the *ΔΔtlo::P_{TET}TLOγ11* strain included those related induced by oxidative stress, "XS_UP", those induced by heavy metal stress, "CU_UP", as well as gene sets related to the proteasome "PROTEASOME COMPLEX (SENSU EUKARYOTA)_CEL", Figure 6.17. The "MED31_UP" set was also enriched in the *ΔΔtlo::P_{TET}TLOγ11* strain. Sets that were enriched comparatively in the *ΔΔtlo* mutant included those containing genes downregulated during oxidative stress, "XS_DN" and those upregulated during infection of RHE cells, "RHE90_UP".

6.3.3 Comparative transcriptomics of strains in YEPD versus under oxidative stress

The transcriptomes of each strain discussed above were also sequenced under oxidative stress conditions in order to determine the transcriptomic response induced by tBOOH and to determine if these responses differed from strain to strain. This analysis was performed because there was a noticeable difference in ability of the strains to tolerate oxidative stress induced by tBOOH in Chapter 5. The AHY940 strain and the *ΔΔtlo::P_{TET}TLOα1* strain were both able to tolerate this stress to similar levels, the *ΔΔtlo* mutant and the *ΔΔtlo::P_{TET}TLOγ11* strain were much less able to tolerate oxidative stress, and the *ΔΔtlo::P_{TET}TLOβ2* strain had an intermediate ability to tolerate oxidative stress induced by tBOOH.

Each strain in YEPD conditions was compared to itself under oxidative stress and lists of differentially expressed genes were generated. The genes that were up- and downregulated are represented in the Volcano plots in Figure 6.18, where the significance of each differential expression call is also highlighted, there were no significantly differentially expressed genes found in the *ΔΔtlo::P_{TET}TLOγ11* strain in tBOOH compared to YEPD. The expression changes in each strain are also represented in a heatmap in Figure 6.19. The expression level of each *TLO* gene in the AHY940 strain was examined and the $\log_2$ fold change in tBOOH compared to YEPD was graphed, Figure 6.20 (A). It was found that in general the expression of the *TLO* genes increased in tBOOH, although the expression of *TLOγ4* and *TLOγ13* decreased in tBOOH compared to in YEPD. When the change in expression of each *TLO* clade was measured, it was found that all clades generally increase in expression also, Figure 6.20 (B).

6.3.3.1 Gene Set Enrichment Analysis

Each list of differentially expressed genes, ranked by $\log_2$ fold change, was subjected to Gene Set Enrichment Analysis to determine if any gene sets were enriched in YEPD conditions or under oxidative stress. In all cases, the "XS_UP" group came back as being highly enriched in oxidative stress conditions, apart from in the case of the *ΔΔtlo::P_{TET}TLOγ11* strain, Figure 6.21. This gene set contains genes that are upregulated upon challenge with xenobiotic stress, specifically oxidative stress. In the *ΔΔtlo::P_{TET}TLOγ11* reintroduced strain it was the case where the "XS_UP" group was enriched in the strain in YEPD rather than under oxidative stress. However, as mentioned there was low reproducibility between replicates in the *ΔΔtlo::P_{TET}TLOγ11* strain, and none of the differentially expressed genes were called with significance.

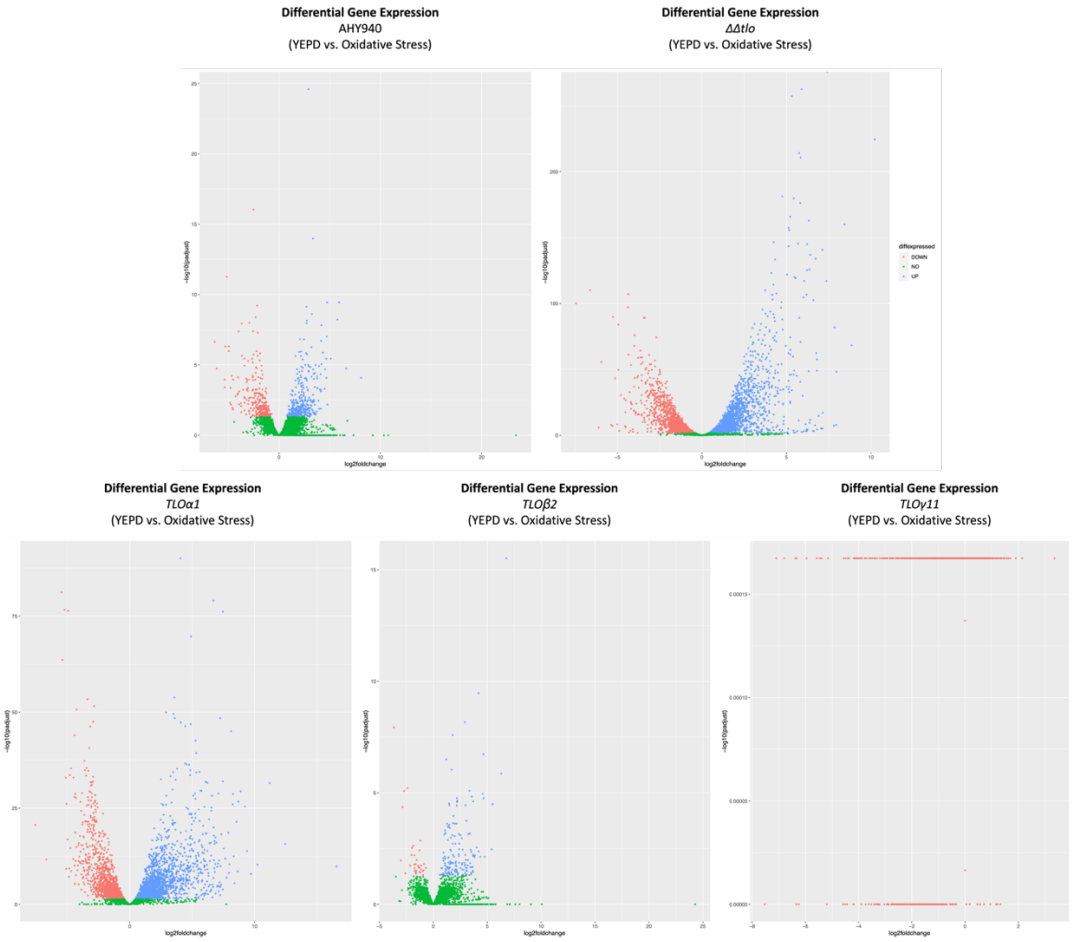


Figure 6.18 Differential gene expression in strains in mid-exponential phase in YEPD at 37 °C versus after exposure to tBOOH for 30 minutes

The $\log_2$ fold change of the expression of each gene is plotted on the x-axis, and the $-\log_{10}$ of the adjusted p value on the y-axis. Points to the right of 0 on the x-axis are expressed more highly in oxidative stress conditions (lower in YEPD), and those to the left are expressed at a lower level in oxidative stress conditions (higher in YEPD). Points in green ("NO") are not significant, those in blue ("UP") are significantly upregulated genes, while red ("DOWN") represents significantly downregulated genes.

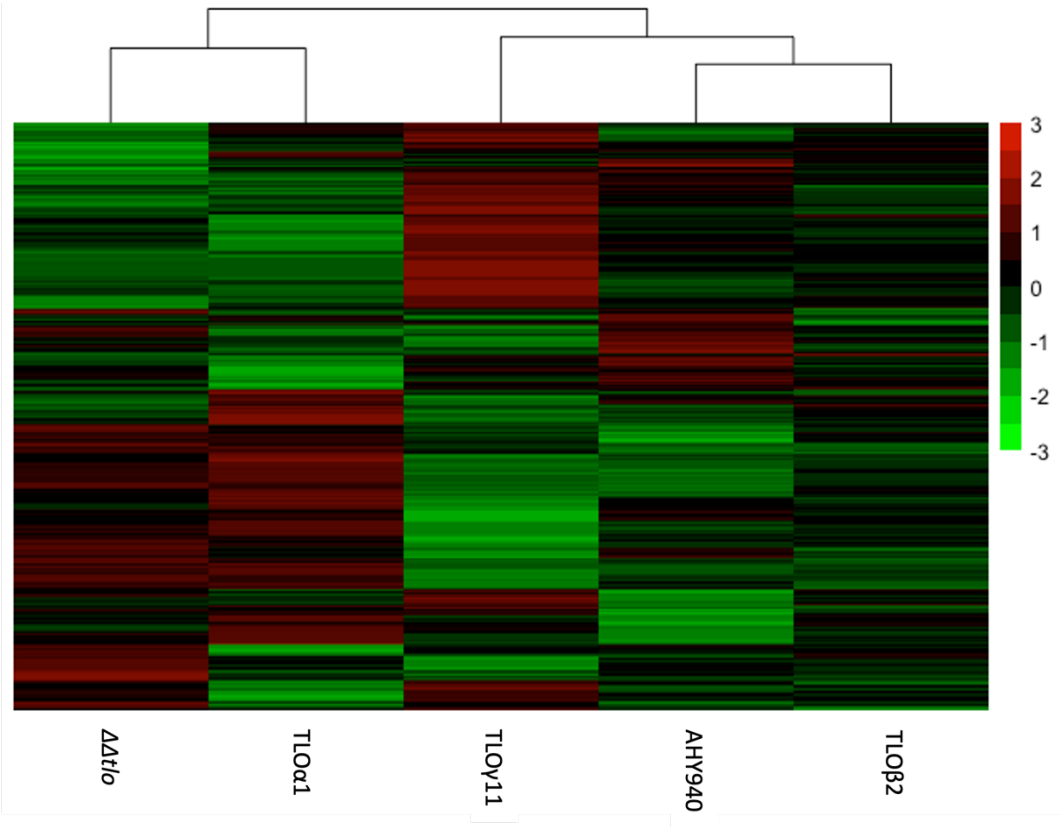


Figure 6.19 Heatmap representing differential expression of genes in tBOOH versus in YEPD

A heatmap was generated from the $\log_2$ fold change values generated by differential gene expression analysis. Genes downregulated in oxidative stress conditions are in green, while genes upregulated are red, see scale bar on right. Hierarchical clustering was performed by default on the strains (top x-axis) to represent similarity in the expression patterns. It should be noted that none of the calls in the *ΔΔtlo::P_{TET}TLOγ11* strain were statistically significant.

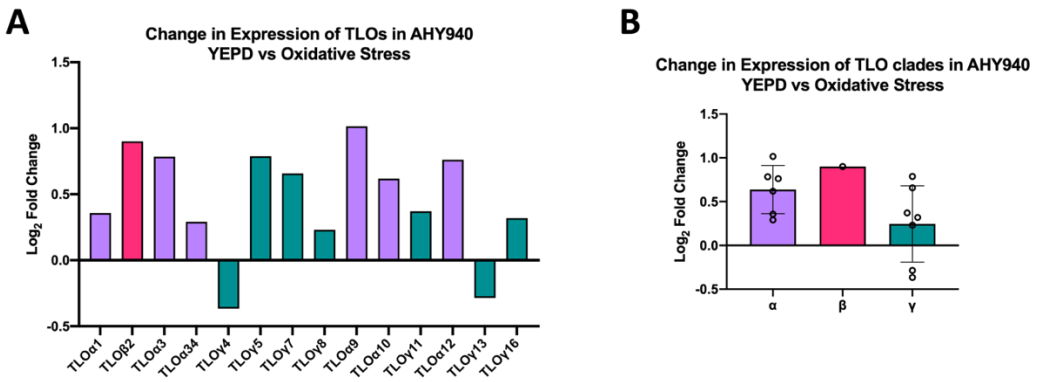


Figure 6.20 Expression change in *TLO* genes in AHY940 upon exposure to oxidative stress

The expression change (log_2 fold change vs. in YEPD) of the *TLO* genes in tBOOH was graphed. (A) Expression change of each individual *TLO*, coloured by clade; α (purple), β (pink), and γ (green), (B) The mean expression change, with standard deviation across the clade, of each clade is shown, with individual points indicated.

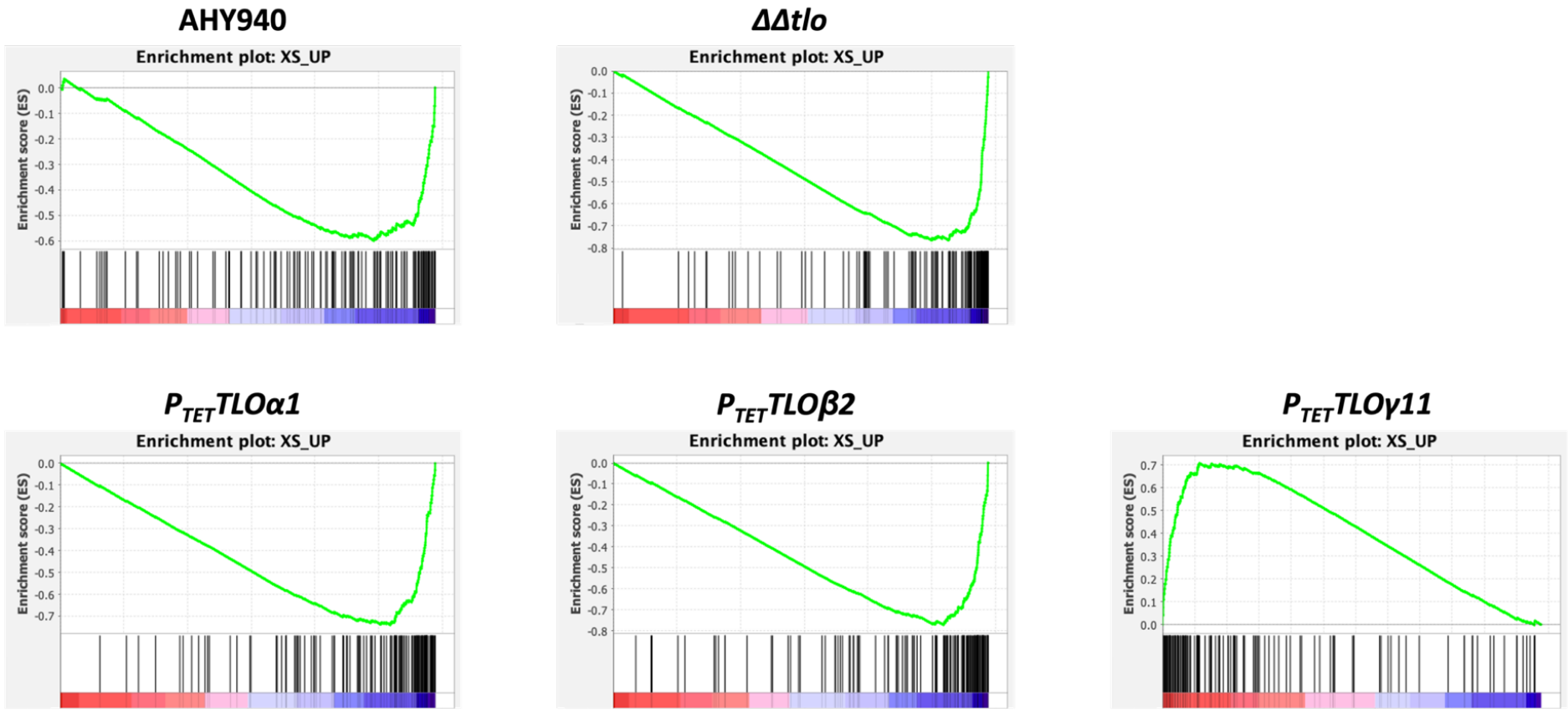


Figure 6.21 Enrichment plots for "XS_UP" set in strains exposed to oxidative stress versus is YEPD

Enrichment plots for the "XS_UP" gene set are shown for each strain above. This set contains genes that are upregulated in response to oxidative stress in *C. albicans*. Enrichment plots are based on the ranked gene lists for each strain were genes are ordered based on their $\log_2$ fold change in tBOOH versus in YEPD. Here, a skew to the right and negative ES values indicate the gene set is enriched in the oxidative stress conditions, skew left means the gene set is more enriched in YEPD. None of the differential gene expression calls for the *ΔΔtlo::PTETTLOγ11* strain were statistically significant.

The fold changes were extracted for each gene in the "XS_UP" gene set, and these were plotted, Figure 6.22 (A), and a heatmap generated, Figure 6.22 (B). This indicated that for all strains, excluding the *ΔΔtlo::P_{TET}TLOγ11*-expressing strain, these genes were upregulated under oxidative stress and were expressed at a higher level in the samples exposed to tBOOH for 30 min than in YEPD. Hierarchical clustering performed by default during generation of the heatmap highlighted that the *ΔΔtlo::P_{TET}TLOγ11* strain was the most different to the others with regards to this response. It also showed that the *ΔΔtlo::P_{TET}TLOα1* and the *ΔΔtlo* mutant exhibited similar transcriptional profiles with regards to these genes. The same can be said for the AHY940 strain and the *ΔΔtlo::P_{TET}TLOβ2* strain.

Similar analysis was also performed on the "XS_DN" gene set genes, Figure 6.23 (A&B), which consists of genes typically downregulated in response to oxidative stress in *C. albicans* cells. This analysis showed that in the cases of the *ΔΔtlo* mutant and the *ΔΔtlo::P_{TET}TLOα1* strain, these genes were downregulated in tBOOH compared to in YEPD. These genes are also downregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain, but not to the same extent. In the case of the AHY940 strain and the *ΔΔtlo::P_{TET}TLOγ11* strain, these genes were upregulated in tBOOH, rather than downregulated. Hierarchical clustering of the columns in the heatmap revealed that the *ΔΔtlo* mutant and the *ΔΔtlo::P_{TET}TLOα1* strain had similar transcript profiles for this gene set, and that the *ΔΔtlo::P_{TET}TLOβ2* and *ΔΔtlo::P_{TET}TLOγ11* strains had similar profiles. The AHY940 strain clustered with the *ΔΔtlo::P_{TET}TLOβ2* and *ΔΔtlo::P_{TET}TLOγ11* strains.

This analysis was also performed using the set of genes that fall under the "Response to oxidative stress" GO term (GO:0006979), Figure 6.24 (A&B). Here it was seen that AHY940, the *ΔΔtlo* mutant, the *ΔΔtlo::P_{TET}TLOα1* strain and the *ΔΔtlo::P_{TET}TLOβ2* strain all showed increased expression of the genes in this set in tBOOH compared to in YEPD. The only strain that did not display this increase was the *ΔΔtlo::P_{TET}TLOγ11* strain, where many of these genes did not change in their expression or were downregulated. Hierarchical clustering of columns in the heat map showed that the transcriptional profile of the *ΔΔtlo::P_{TET}TLOγ11* strain was distinctly different from all other strains. The *ΔΔtlo* mutant and the *ΔΔtlo::P_{TET}TLOα1* strain clustered together, and the AHY940 and the *ΔΔtlo::P_{TET}TLOβ2* strain clustered together.

Some differentially expressed genes differed between strains included *ERG11, SSU81* and *HSP12*. The. *ERG11* gene is not significantly differentially expressed in AHY940 or

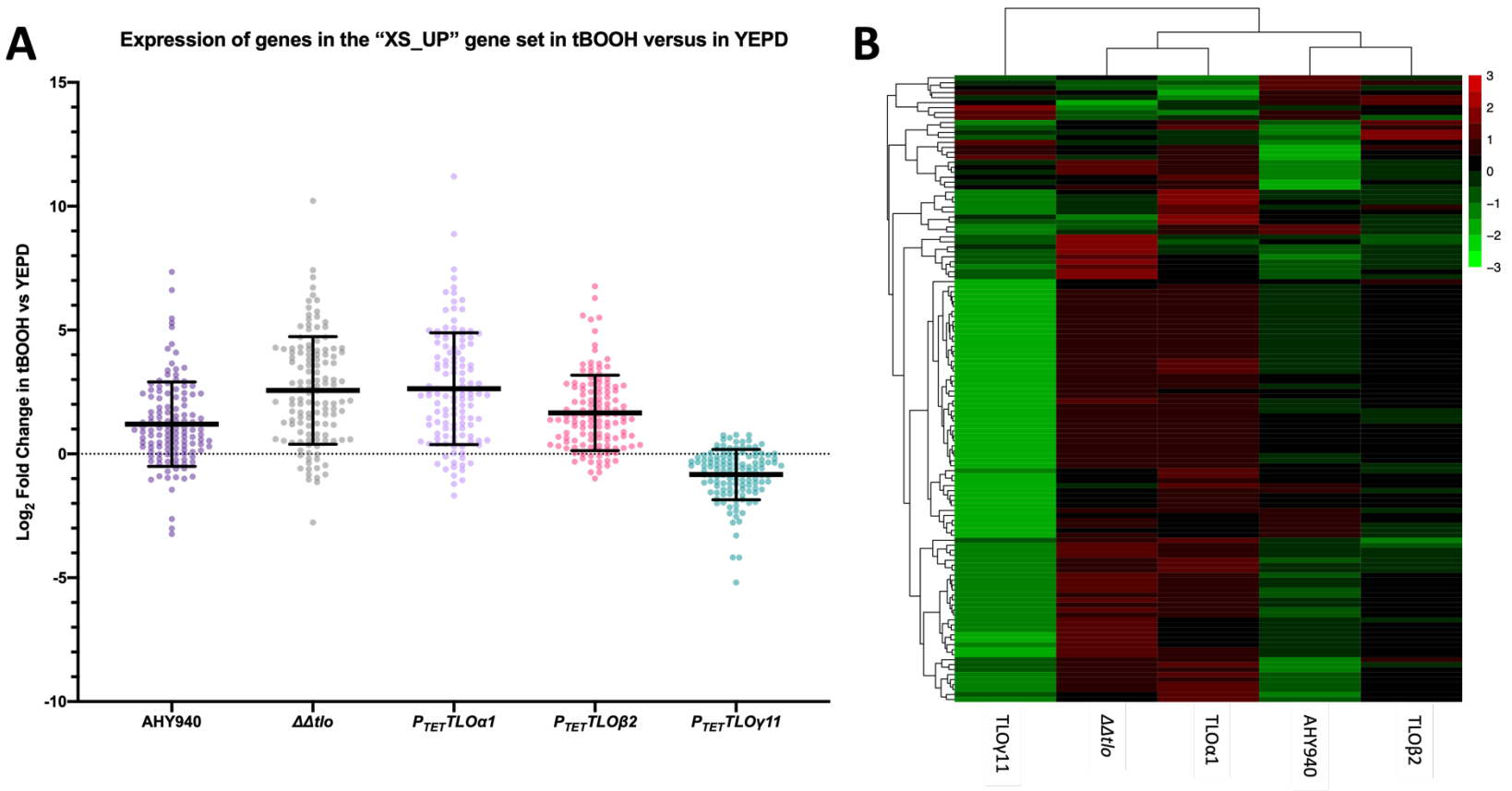


Figure 6.22 Expression of the "XS_UP" gene set in strains exposed to oxidative stress versus in YEPD

(A) Expression values (log_2 fold change in tBOOH vs. in YEPD) were plotted for each gene in the "XS_UP" set (genes upregulated in response to oxidative stress) for each of the strains; AHY940 (dark purple), *ΔΔtlo* (grey), *ΔΔtlo::P_{TET}TLOα1* (light purple), *ΔΔtlo::P_{TET}TLOβ2* (pink), and *ΔΔtlo::P_{TET}TLOγ11* (green). (B) Heatmap representing the log_2 fold change of each gene in the set for strains in tBOOH vs. in YEPD. Green indicates downregulation in tBOOH and red indicates upregulation in tBOOH. Hierarchical clustering indicates similarity between responses.

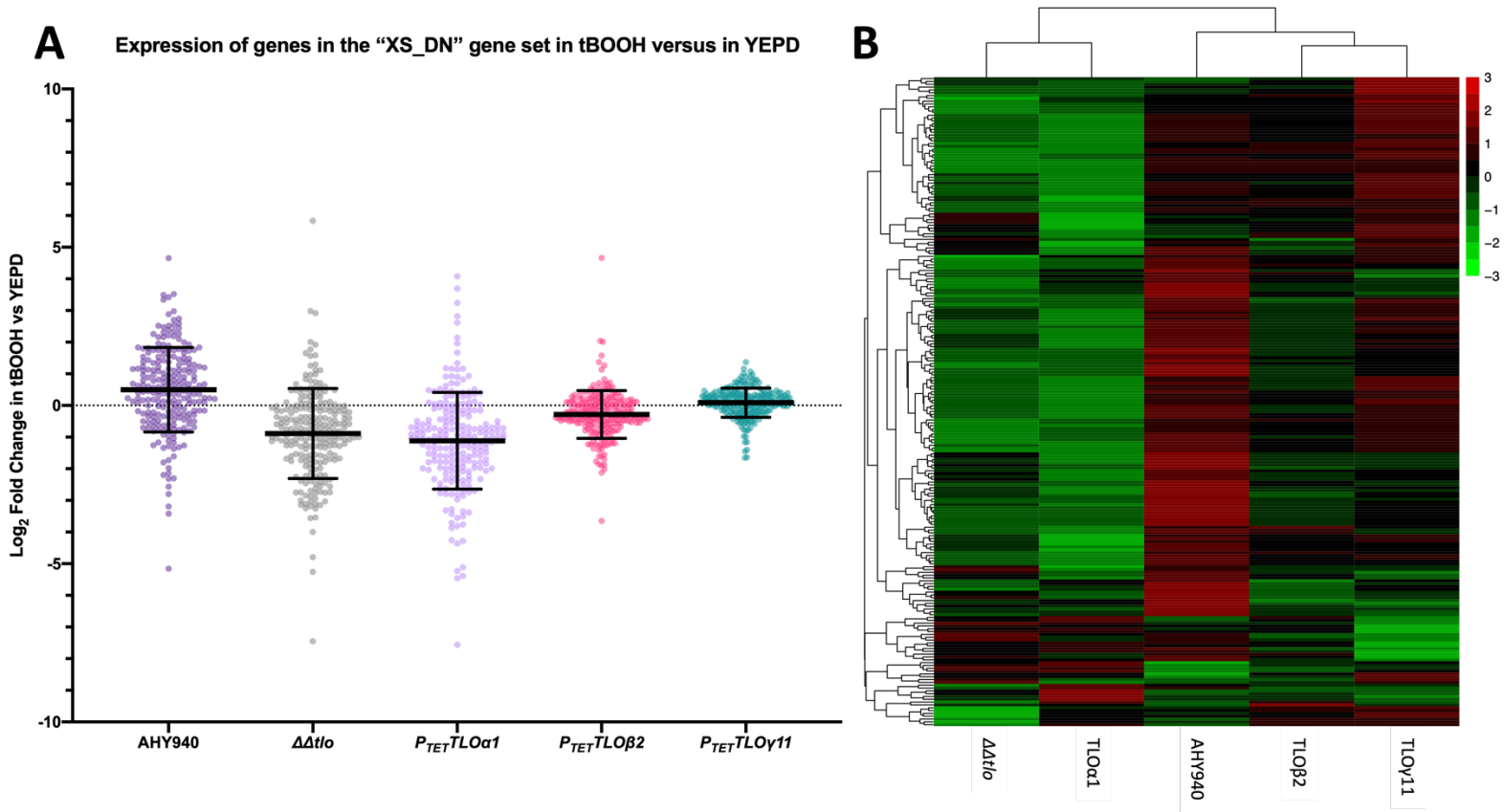


Figure 6.23 Expression of the "XS_DN" gene set in strains exposed to oxidative stress versus in YEPD

(A) Expression values (log_2 fold change in tBOOH vs. in YEPD) were plotted for each gene in the "XS_DN" set (genes downregulated in response to oxidative stress) for each of the strains; AHY940 (dark purple), *ΔΔtlo* (grey), *ΔΔtlo::P_{TET}TLOα1* (light purple), *ΔΔtlo::P_{TET}TLOβ2* (pink), and *ΔΔtlo::P_{TET}TLOγ11* (green). (B) Heatmap representing the log_2 fold change of each gene in the set for strains in tBOOH vs. in YEPD. Green indicates downregulation in tBOOH and red indicates upregulation in tBOOH. Hierarchical clustering indicates similarity between responses.

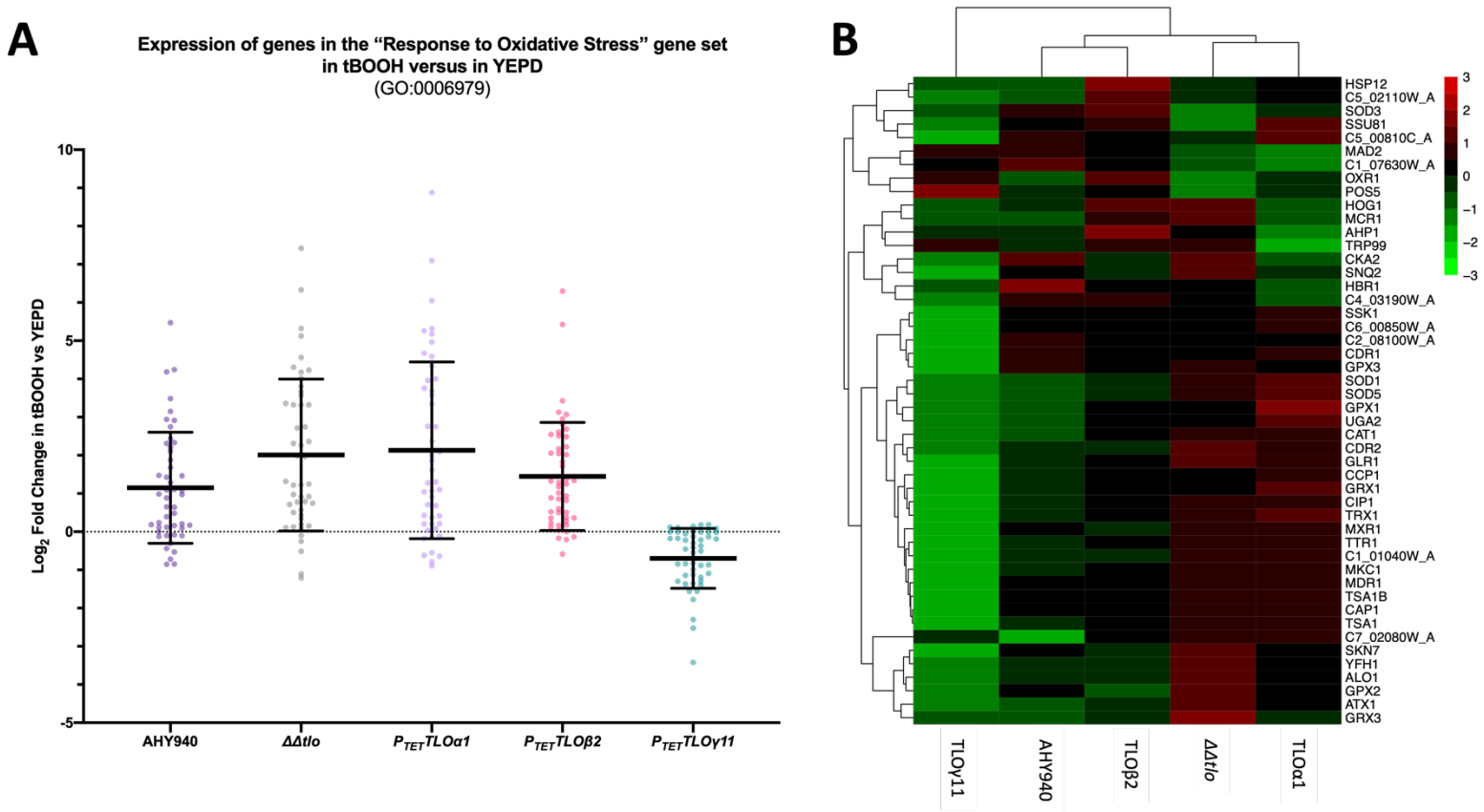


Figure 6.24 Expression of the "Response to Oxidative Stress" gene set in strains exposed to oxidative stress versus in YEPD

(A) Expression values (log_2 fold change in tBOOH vs. in YEPD) were plotted for each gene in the "Response to Oxidative Stress" (GO:0006979) set for each of the strains; AHY940 (dark purple), *ΔΔtlo* (grey), *ΔΔtlo::P$_{TET}$TLOα1* (light purple), *ΔΔtlo::P$_{TET}$TLOβ2* (pink), and *ΔΔtlo::P$_{TET}$TLOγ11* (green). (B) Heatmap representing the log_2 fold change of each gene in the set for strains in tBOOH vs. in YEPD. Green indicates downregulation in tBOOH and red indicates upregulation in tBOOH. Gene names are labelled on the right. Hierarchal clustering indicates similarity between responses.

the *ΔΔtlo::P_{TET}TLOβ2* strain in YEPD compared to in tBOOH, however this gene is significantly downregulated in tBOOH in the *ΔΔtlo* mutant and in *ΔΔtlo::P_{TET}TLOα1* (-1.2 and -1.8 L_2FC in tBOOH vs. YEPD respectively). The *SSU81* gene was significantly upregulated in AHY940 and in the *ΔΔtlo::P_{TET}TLOα1* strain in tBOOH (+1.3 and +1.9 L_2FC respectively), but not significantly different in the *ΔΔtlo* mutant or the *ΔΔtlo::P_{TET}TLOβ2* strain. The *HSP12* gene was not significantly differentially expressed in AHY940 or the *ΔΔtlo* mutant, however this gene was significantly upregulated in the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains in tBOOH (+0.9 and +2.2 L_2FC respectively).

The transcriptomes of the strains grown in tBOOH were also compared to each other to examine differential expression of genes between strains, data not shown. Gene set enrichment analysis showed that compared to AHY940, the *ΔΔtlo* mutant was expressing typically upregulated during oxidative stress more, while the AHY940 strain was expressing those typically downregulated during oxidative stress. The *ΔΔtlo* mutant strain was also enriched for genes typically upregulated following exposure to heavy metals, and after phagocytosis by host immune cells.

Compared to the *ΔΔtlo::P_{TET}TLOα1* strain, the *ΔΔtlo* mutant was enriched for genes typically upregulated during oxidative stress. Comparatively, the *ΔΔtlo::P_{TET}TLOα1* strain was enriched for genes typically downregulated during oxidative stress, and downregulated during heat shock. In comparison to the *ΔΔtlo::P_{TET}TLOβ2,* the *ΔΔtlo* mutant was enriched for genes typically upregulated during oxidative stress, and the *ΔΔtlo::P_{TET}TLOβ2* strain was enriched for genes downregulated during oxidative stress. Comparison to the *ΔΔtlo::P_{TET}TLOγ11* strain also highlighted that the *ΔΔtlo* mutant was enriched for genes typically upregulated during oxidative stress. This data indicated that the *ΔΔtlo* mutant was expressing the genes typically upregulated upon exposure to oxidative stress at the highest level compared to all other strains tested.

Comparisons to the AHY940 transcriptome highlighted that the *ΔΔtlo::P_{TET}TLOα1* strain was enriched for genes typically upregulated during oxidative stress. Comparison to the AHY940 transcriptome highlighted that the *ΔΔtlo::P_{TET}TLOβ2* strain was also enriched for the genes that are upregulated upon exposure to oxidative stress. Comparison of AHY940 and the *ΔΔtlo::P_{TET}TLOγ11* strain did not highlight enrichment for any gene sets related to the oxidative stress response, but did highlight that the "WHITE_UP" and "MED31_DN" gene sets were upregulated in the AHY940 strain in comparison.

6.4 Discussion

The RNA sequencing approaches used in this chapter facilitated investigation into the differences in gene expression in the mutant and *TLO*-expressing strains in both nutrient rich culture conditions, and under oxidative stress. This enabled deeper investigation into the role that the *TLO* gene family may play in the *C. albicans* cell, and what effect individual members may have on transcription. A summary diagram of the findings from this chapter can be found in Figure 6.25, where the main gene sets up and downregulated in each strain compared to the *ΔΔtlo* mutant (in YEPD) are highlighted.

6.4.1 Transcriptomic comparison of the ΔΔtlo *mutant and AHY940*

The comparison of the *ΔΔtlo* mutant and the AHY940 parent strain was performed to determine what effect deletion of the entire gene family would have on the transcriptome of *C. albicans*. This comparison also served as further confirmation that all *TLO* genes had been effectively deleted and that there was no expression of *TLO* genes in the *ΔΔtlo* mutant.

Examination of the genes that were upregulated in the *ΔΔtlo* mutant found that these genes had mainly had roles in ribosome biogenesis and assembly. The upregulation of these genes was seen in both the GO analysis of the significantly upregulated genes in the *ΔΔtlo* mutant, as well as GSEA of the ranked list of all differentially expressed genes between the *ΔΔtlo* mutant and AHY940. The GSEA analysis indicated that the *ΔΔtlo* mutant strain was enriched for gene sets encoding proteins that interact with proteins such as Nip7, Has1, Rlp7, and Rrp5, which are all involved in ribosome biogenesis. The ribosome is a complex made up of ribosomal RNA and protein that functions to convert messenger RNA into polypeptide chains, and the complex is highly conserved throughout eukaryotes. Biogenesis of the ribosome is tightly regulated and requires a large amount of energy from the cell, for synthesis and trafficking of these components, with yeast cells making up to 2,000 ribosomes per minute in the exponential phase of growth (Woolford and Baserga, 2013). Dysregulation of ribosome biosynthesis can have major deleterious consequences for a cell, and while most of the literature focuses on downregulation and loss of function of the ribosome (Woolford and Baserga, 2013), the uncontrolled upregulation of ribosome biosynthesis could also wreak havoc in a cell, with a possible consequence including the mismanagement of cellular resources

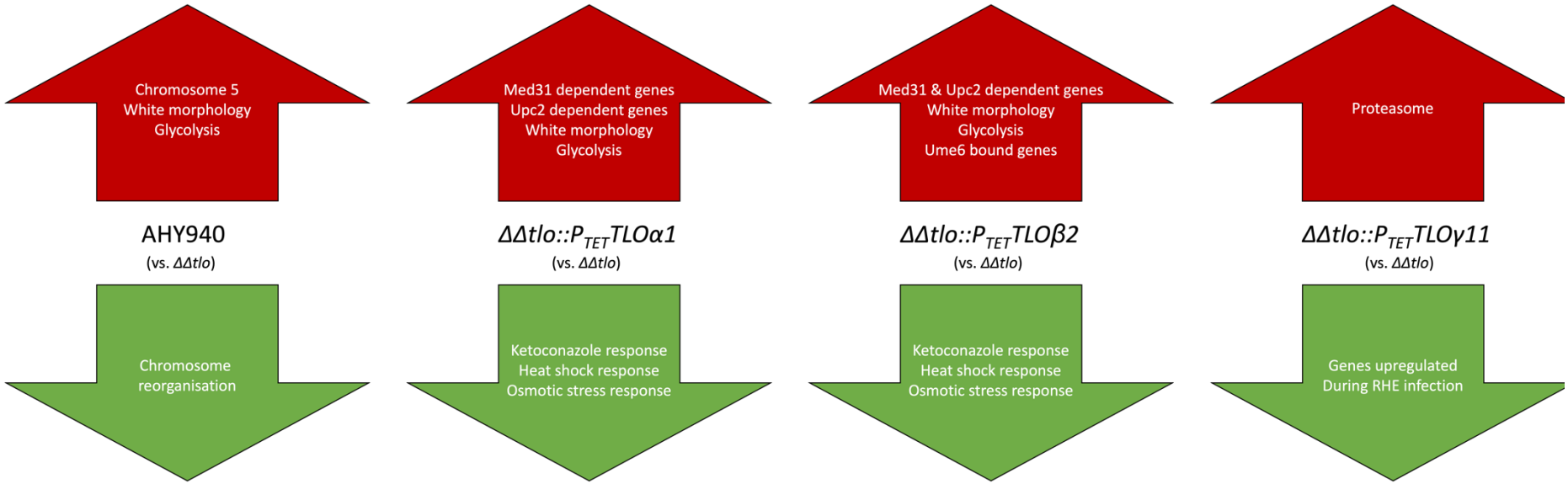


Figure 6.25 Summary of gene sets up and downregulated in *TLO* mutant strains

The findings from the transcriptomic analysis performed in this chapter is found above. Gene sets that were upregulated in *TLO* mutant strains (or WT AHY940) compared to the *ΔΔtlo* mutant when strains were grown to mid-exponential phase in YEPD are listed in the red arrows (pointing up). Those that are downregulated are listed in the green arrows (pointing down).

resulting in loss of fitness. The increase in ribosome formation could also an attempt to synthesis proteins in response to the poor metabolic growth seen in this strain.

Some other groups of genes that were found to be upregulated in the *ΔΔtlo* mutant compared to AHY940 were those related to chromosome organisation and DNA recombination. It has already been highlighted in Chapter 4 that the *ΔΔtlo* mutant displayed an atypical chromosomal banding pattern when run on a CHEF gel, indicating the possibility of chromosomal rearrangements having occurred. Pairing these data together indicates a possible role for the *TLO* genes in maintaining chromosomal stability in *C. albicans*. Specifically upregulated were the *MRE11* and *RAD50* genes, which encode subunits of the Mre11-Rad50-Xrs2 complex involved in repairing double strand DNA breaks. This complex is not well defined in *Candida*, however, in *S. cerevisiae* it has been shown that this complex is not necessary for viability as it is in mammalian cells (Rupnik *et al.*, 2008). In mammalian cells the MRN complex (the Xrs2 protein in yeast is homologous to the Nbs1 protein in mammalian cells) has also been shown to regulate homologous recombination, not just in response to DNA damage, but throughout the entire cell cycle (Buis *et al.*, 2012). The *ΔΔtlo* mutant could be eliciting a persistent and unwarranted DNA damage response, and upregulating molecules responsible for recombination such that the chromosomal structure of the mutant is being altered.

The enrichment of the glycolysis gene set in AHY940 compared to the *ΔΔtlo* mutant appears to be consistent with the phenotypes of these strains described in Chapter 4; i.e. that the growth rate of the *ΔΔtlo* mutant is much slower than that of AHY940 in YEPD. The transcriptomic data show that the genes involved in glycolysis are being expressed at a much lower level in the *ΔΔtlo* mutant than in AHY940. These include genes coding enzymes involved in the glycolytic pathway such as, *ADH1, CDC19, ENO1, GPM1* and *PFK1&2*, as well as one of the transcriptional regulators that control the pathway, *TYE7* (Askew *et al.*, 2009). The lower level of expression of these genes appears to have a direct effect on the ability of the *ΔΔtlo* mutant to grow in YEPD, and points towards a role for the *TLO* family in affecting transcription of glycolytic genes. A *Δtlo1/Δtlo2* mutant *C. dubliniensis* strain also showed downregulation of many genes in the glycolytic pathway, including *ENO1 FBA1, GPM1 PFK1* and *PFK2* as well as the *TYE7* transcriptional regulator. This *Δtlo1/Δtlo2* mutant was also much slower growing in YEPD than its WT *C. dubliniensis* parent (Haran *et al.*, 2014). These data point to a role for the *TLO*s in controlling the expression of genes involved in the carbohydrate metabolism pathways, directly impacting the ability of the cell two properly utilise

carbon and grow in nutrient rich conditions, ultimately contributing to the overall fitness of the cell.

The AHY940 strain also shows enrichment for the "WHITE_UP" group of genes, which are genes typically expressed in the white cell phase. Typical *C. albicans* blastospores are white cells, displaying the typical round/oval cell morphology, however under certain conditions, an epigenetic change occurs and the cells switch to the opaque morphology, which is more elongated, rod shaped, and is the mating competent cell type (Takagi *et al.*, 2019). In Chapter 4 the morphology of the AHY940 strain and the *ΔΔtlo* mutant were compared, and it was seen that the AHY940 strain presents as typical yeast blastospores, while the *ΔΔtlo* mutant grows pseudohyphally. The enrichment of the "WHITE_UP" gene set in the AHY940 strain indicates that there are white phase specific genes that are not being highly expressed in the *ΔΔtlo* mutant, which may be contributing to the irregular morphology of this strain.

It was previously highlighted that the AHY940 strain contained a trisomy with respect to chromosome 5. This trisomy was reflected in the gene expression profile of this strain. In comparison to the *ΔΔtlo* mutant, gene sets containing genes found on chromosome 5 were highlighted by GSEA as being significantly enriched in the AHY940 strain. Gene sets covering the entire length of chromosome 5 are enriched in AHY940, indicating a complete extra copy of the chromosome possibly being present in this strain. The increased expression of genes from chromosome 5 by the AHY940 strain also means that the rest of the analysis involving this strain may be skewed, particularly for genes on chromosome 5. It appears to be the case that the AHY940 strain is enriched for gene sets that are increased by ketoconazole exposure. The *ERG11* gene product is the target of azole drugs, it is a member of both the "KETOCONAZOLE_UP" gene set and the "Ca21Chr5x1" gene set (1-250,000 bp), and increased mRNA levels of this gene have been associated with azole resistance. Duplication of the left arm of chromosome 5 has been shown to increase resistance of *C. albicans* strains to azoles, specifically by increasing the expression of *ERG11* (Selmecki *et al.*, 2008). It would be of interest to compare the resistance of AHY940 to azole drugs to that of WT *C. albicans* SC5314, which does not have the trisomy of chromosome 5, to determine if this trisomy is affecting the ability of the AHY940 strain to resist azole drugs. A new *Δleu2/LEU2* heterozygous derivative of SC5314 has been constructed, and the experiments performed in this work will be repeated with this strain, and azole susceptibility testing

will also be performed to compare AHY940, this new mutant and the parent SC5314 strain.

The increased expression of genes on chromosome 7 in the *ΔΔtlo* mutant compared to AHY940 could also indicate further issues with the karyotype of AHY940, possibly a monosomy of chromosome 7 (or a trisomy of this chromosome in the *ΔΔtlo* mutant), although this was not highlighted by analysis performed by the Anderson lab (personal communication). In all, the trisomy of chromosome 5 in the AHY940 strain appears to confound the comparison between these strains, and so further comparisons were only drawn between the *TLO* reintroduction strains and the *ΔΔtlo* mutant, though in many cases the results are similar.

6.4.2 Transcriptomic comparison of TLO *reintroduction strains and the* ΔΔtlo *mutant*

Comparison of the *TLO* reintroduction strains in Chapter 5 revealed that by expressing *TLOα1* in the *ΔΔtlo* mutant background, many phenotypes could be restored, and in many cases the *TLOα1* reintroduction strains were indistinguishable from the AHY940 WT. Expression of the *TLOβ2* gene in the *ΔΔtlo* mutant background could also restore many phenotypes but did also have the effect of causing constitutive hyphal growth in a subpopulation of cells of this strains The expression of *TLOγ11* gene in the *ΔΔtlo* mutant background did not appear to have much effect on phenotype, with these strains performing similarly to the *ΔΔtlo* mutant in the phenotypic assays performed.
In order to investigate if individual *TLO* genes have specific functions, RNA was extracted from the strains expressing the three clade-representative *TLO* genes under the *TET1* promoter (the weaker of the two tested) during mid-exponential growth in YEPD and sequenced to investigate what transcriptomic effects *TLO* expression in the *ΔΔtlo* mutant background may have. One of the first observations from these experiments was that the expression of *TLO* genes from the reintroduced constructs in the *ΔΔtlo* background was detected. This had previously been determined by qRT PCR as detailed in Chapter 5, and these experiments confirmed those results. Comparison with the expression of each single *TLO* gene in the AHY940 strain compared to the *ΔΔtlo* mutant showed that in the reintroduction strains, under the *TET1* promoter, the *TLO*s were being expressed at a lower level than in the WT (which of course could have up to 14 different TLO genes being expressed at any given time). As described in the results section, the *ΔΔtlo::*P_{TET}*TLOγ11* strain showed increased expression of several *TLO* transcripts, however, it appears to be the case that the similarity of the genes may have

resulted in the misattribution of transcripts from the *ΔΔtlo::P_{TET}TLOγ11* cassette to other *TLO*s, specifically *TLOγ4, TLOγ16* and to a lesser extent *TLOα9*. This issue with RNA-seq for measuring the expression from *TLO* gene family members has been mentioned previously by Anderson *et al.* in 2012, where they postulate that the ambiguous mapping of highly similar transcripts may mask the biological reality of gene expression (Anderson *et al.*, 2012). They suggest that a better approach is the use of gene specific oligonucleotide primers to measure the expression by qRT PCR. The fact that this has also been done as part of this study has confirmed the expression from the cassette, but also confirms this ambiguous mapping phenomenon.

Comparison of the differential gene expression in the *TLO* reintroduction strains compared to the *ΔΔtlo* mutant showed that the expression of each *TLO* in the *ΔΔtlo* background had overlapping as well as some unique effects on the transcriptome, even in the case of the *ΔΔtlo::P_{TET}TLOγ11* strain, which seems to have a very limited effect on phenotype, at least under the conditions studied. Differential expression analysis comparing the transcriptomes of the reintroduction strains to AHY940 was not performed as part of this work due to the chromosome 5 trisomy present in AHY940. However, this analysis will be performed when a new AHY940 strain (*Δleu2/LEU2* derivative of SC5314) has been constructed.

Analysis of differential gene expression in the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains compared to the *ΔΔtlo* mutant showed that the expression of these *TLO*s significantly affected expression of a considerable number of genes, and of these genes, many of the same genes were up or downregulated by both Tlo proteins. Gene set enrichment analysis of the ranked lists of differentially expressed genes in each of these strains compared to the *ΔΔtlo* mutant also highlighted similar sets of genes that were enriched in these two *TLO* reintroduction strains compared to the *ΔΔtlo* mutant. The top enrichments for both *TLO* reintroduction strains were the groups "KETOCONAZOLE_DN" and "UPC2_DN", and conversely, the top enrichments in the *ΔΔtlo* mutant strain compared to these strains were the "KETOCONAZOLE_UP" and "UPC2_UP" groups. This is the opposite of what was observed in the AHY940 comparison, however, as mentioned, the trisomy of chromosome 5 complicates the interpretation of these results. These gene sets are related, and contain many of the same genes. Of the 151 genes in the "KETOCONAZOLE_DN" set, 45 of them (~30%) are also found in the "UPC2_DN" set. For the "KETOCONAZOLE_UP" set, 95 of the 200 genes (~48%) in the set are also found in the "UPC2_UP" set.

Ketoconazole exposure increases the expression of genes involved in ergosterol biosynthesis, the *ERG* family of genes involved in ergosterol biosynthesis, including the gene that encodes the target of azole drugs, *ERG11* (Liu *et al.*, 2005). However, *ERG11* is downregulated in the *ΔΔtlo* mutant compared to the *ΔΔtlo::*P_{TET}*TLOα1* and *ΔΔtlo::*P_{TET}*TLOβ2* reintroduction strains. While *ERG11* is a member of the "KETOCONAZOLE_UP" gene set its expression in the *ΔΔtlo* mutant does not contribute to the enrichment of the gene set in the GSEA, i.e. other genes are contributing to the enrichment of the set. *UPC2* encodes a transcription factor that controls the expression of genes in the ergosterol biosynthesis pathway, including *ERG11* (Vasicek *et al.*, 2014). *UPC2* is also downregulated in the *ΔΔtlo* mutant. The enrichment for the "UPC2_UP" set in the *ΔΔtlo* mutant indicates that genes that are repressed by Upc2 are upregulated, suggesting that Upc2 requires Tlo for repression activity. Conversely, the enrichment for "UPC2_DN" gene set in the *TLO* reintegrated strains indicates increased expression of genes that require Upc2 for expression, suggesting that Tlo is required for Upc2 mediated activation. Work done by a postdoctoral researcher in the lab, Dr James O'Connor Moneley, found that the *ΔΔtlo* mutant strain is more tolerant of azole drugs (fluconazole and posaconazole) than the WT AHY940 and the *ΔΔtlo::*P_{TET}*TLOα1*- and *ΔΔtlo::*P_{TET}*TLOβ2*-expressing strains. Fluconazole acts by targeting the product of the *ERG11* gene, lanosterol 14α-demethylase, an enzyme in the ergosterol biosynthesis pathway, and inhibiting it. This disruption leads to a build-up of toxic intermediates in the precursor stages of ergosterol biosynthesis, inhibiting growth of the cell (Vasicek *et al.*, 2014). Analysis performed by Dr O'Connor Moneley with collaborators in Swansea University examined the sterol content of the *ΔΔtlo* mutant cell membrane and found that this strain already contains significantly lower levels of ergosterol than WT strains, and significantly higher levels of toxic intermediates. Perhaps the absence of a target for the azole drugs, as shown by the downregulation of *ERG11* seen in this RNA-seq analysis and the finding that there is decreased ergosterol in the *ΔΔtlo* mutant, allows for increased tolerance in the *ΔΔtlo* mutant. As the Upc2 transcription factor itself is an activator of genes in the ergosterol biosynthesis pathway, including *ERG11,* the downregulation of *UPC2* in the absence of *TLO*s in the *ΔΔtlo* mutant could be the reason the *ERG* genes are not being expressed, resulting in the aberrant composition of the cell membrane.

There is also evidence that deletion of *ERG11* in *C. albicans* leads to increased sensitivity to oxidative stress, cell wall perturbing compounds such as Calcofluor White and Congo

Red, as well as defective filamentation, all of which are phenotypes that have been observed in the *ΔΔtlo* mutant (see Chapter 4). The *ΔΔerg11* mutant was also avirulent in a mouse model (Wu *et al.*, 2018). The azole tolerant phenotype has also been found to be reversed by the expression of *TLOα1* or *TLOβ2* in the null mutant background, but not the by the expression of *TLOγ11,* an indication that *TLOα1* and *TLOβ2* may play a role in controlling ergosterol biosynthesis in *C. albicans*. Further research on this azole tolerance phenotype in the *ΔΔtlo* mutant is being conducted to investigate the role of the *TLO* gene family in this phenomenon.

The GSEA of the *ΔΔtlo::P_{TET}TLOα1* strain compared to the *ΔΔtlo* mutant highlighted enrichment of genes found to interact with Erg3, Erg5 and Erg28 in *S. cerevisiae*. This enrichment was not found in the *ΔΔtlo::P_{TET}TLOβ2* strain compared to the *ΔΔtlo* mutant, indicating a possible difference in the presence or activity of these proteins. These Erg proteins are also involved in the ergosterol biosynthesis pathways. Differences in the tolerance of the *ΔΔtlo::P_{ENO}TLOα1* and *ΔΔtlo::P_{ENO}TLOβ2* strains to azole drugs was also detected in the work performed by Dr O'Connor Moneley in the lab. This remains to be tested in the P_{TET} strains, however, this data paired with the RNA-seq data here may indicate differences in the sterol content of the cell membranes of the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains resulting in different levels of drug susceptibility.

The GSEA comparison of the *TLO* reintroduction strains to the *ΔΔtlo* mutant also highlighted that gene sets related to the opaque cell type were enriched in the *ΔΔtlo* mutant compared to the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains, but not compared to the *ΔΔtlo::P_{TET}TLOγ11* strain. This data also conforms with the finding that the "WHITE_UP" gene set was significantly enriched in AHY940 compared to the *ΔΔtlo* mutant. White and opaque cells are two distinct cell types that *C. albicans* can present as, described in Section 6.4.1. As previously mentioned, the morphology of AHY940 and the *ΔΔtlo::P_{TET}TLOα1* cells are typical yeast cell morphology, indicating that these are likely to be white cells. In the *ΔΔtlo::P_{TET}TLOβ2* strain a small percentage of cells do form true hyphae, however the rest of the cells grow also grow in the typical yeast morphology, again, apparently exhibiting typical White cell morphology. The *ΔΔtlo* mutant and the *ΔΔtlo::P_{TET}TLOγ11* strain grow constitutively as pseudohyphae, however any cells that were separate to pseudohyphal masses had irregular cellular morphologies. It could be the case that these strains are in fact growing as opaque-like cells. Members of the lab are also investigating this in more depth, specifically using media supplemented with phloxine B dye to selectively stain opaque colonies, and

attempting to fluorescently tag opaque specific genes, such as *OP4, WOR1* and *EFG1,* to examine their expression in the cells (Xie *et al.*, 2013). However, to date there is no evidence that deleting the *TLO* gene family affects the rate of white-opaque switching.

The *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains, like AHY940, can grow at a much faster rate in YEPD than the *ΔΔtlo* mutant, indicating that the expression of *TLOα1* and *TLOβ2* in an otherwise *TLO* free background can restore efficient carbohydrate metabolism in these strains. The expression of *TLOγ11* in the *ΔΔtlo* mutant did not restore growth rate in YEPD. The GSEA highlighted that the glycolysis gene set was enriched in both *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* compared to the *ΔΔtlo* mutant. This set was not found to be significantly enriched in the *ΔΔtlo::P_{TET}TLOγ11* or the *ΔΔtlo* mutant in that comparison, indicating that there is no significant difference in the expression of these genes in either strain. In the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains increased expression of the transcription factor *TYE7*, which controls the expression of many genes downstream in the glycolytic pathway, was evident (Askew *et al.*, 2009). Some of glycolysis genes were downregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain compared to the *ΔΔtlo* mutant, such as *PGI1, PFK26* and *TPI1*, which did not appear to affect the growth rate in YEPD as described in Chapter 5. Genes *PGI1* and *TPI1* are under the control of the Gal4 transcription factor, which is another activator of genes involved in glycolysis, but has been found to activate glycolytic genes to a lesser extent than Tye7 (Askew *et al.*, 2009).

The expression of *GAL4* was not found to be significantly different in the *ΔΔtlo* mutant than in either of the *ΔΔtlo::P_{TET}TLOα1* or *ΔΔtlo::P_{TET}TLOβ2* strains, however Gal4 has only been reported to be important in fermentative carbon utilisation, and that in respiratory carbon use the expression of *TYE7* can compensate for the absence of *GAL4* (Askew *et al.*, 2009). There does appear to be a difference in expression of genes regulated by *GAL4,* namely *PGI1* and *TPI1*, which are expressed at a lower level in the *ΔΔtlo::P_{TET}TLOβ2* strain than in the *ΔΔtlo::P_{TET}TLOα1* strain. While the downregulation of these genes does not appear to have any effect on the growth rate of the *ΔΔtlo::P_{TET}TLOβ2* strain in YEPD (with 2% glucose), deficits may become apparent if grown in fermentative conditions. Without the explicit downregulation of *GAL4* in the *ΔΔtlo::P_{TET}TLOβ2* strain compared to the other strains, but with the downregulation of genes controlled by it, it may be possible that Tloβ2 itself is repressing the expression of the genes downstream of *GAL4*. In the GSEA, the gene sets bound by both Tye7 and Gal4 were significantly enriched in only the *ΔΔtlo::P_{TET}TLOα1* strain compared to the

ΔΔtlo mutant, and not in the *ΔΔtlo::P_{TET}TLOβ2* strain. However the gene set bound by Tye7 fell just outside the significance cut off in the *ΔΔtlo::P_{TET}TLOβ2* strain indicating that there was enrichment of these genes in this strain compared to the *ΔΔtlo* mutant.

The promoter regions of the telomeric *TLO* family members have been shown to have Gal4 binding domains within them, and it had previously been suggested that expression of the *TLO* genes could be under some level of control by Gal4 (Sullivan *et al.*, 2015). However, as the *TLO* genes in this chapter are under the control of the *TET1* promoter, and there are no other *TLO*s present under their native promoters, this is not the model to study Gal4 regulation on *TLO*s.

Gene ontology analysis of genes downregulated in the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains compared to the *ΔΔtlo* found the enrichment of many genes related to autophagy. The *ΔΔtlo* mutant strain was found to be expressing many of these genes at a high level compared to AHY940, and it was thought to be a cellular response to the strain's inability to utilise external nutrients, e.g. through carbon metabolism. The *ΔΔtlo* mutant strain exhibits a constitutive starvation response, even in the presence of nutrients, similar to the transcriptome of a *C. dubliniensis Δtlo1/Δtlo2* mutant (Haran *et al.*, 2014). The reintroduction of *TLOα1* and *TLOβ2* did not only upregulate genes required for nutrient utilisation, it also downregulated the genes involved in autophagy recycling, restoring cellular homeostasis.

Gene sets that were found to be enriched in the *ΔΔtlo* mutant compared to both the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains included sets of genes typically upregulated upon exposure to and engulfment by, macrophages, such as "BMDM_PHAGOCYTOSIS_UP" and "RHE90_UP". These sets were also found to be enriched in a *C. dubliniensis Δtlo1/Δtlo2* mutant when compared to its *C. dubliniensis* WT parent (Haran *et al.*, 2014). Enrichment of these gene sets indicate a role for the *TLO* genes during the infection of host tissue and possibly a role in invading the host immune response. Expression of different *TLO* genes could be involved in triggering the switch from commensal to pathogen upon sensing of the right conditions. These gene sets were also enriched in the *ΔΔtlo* mutant compared to AHY940, and this could be indicative of the *ΔΔtlo* mutant undergoing a constitutive stress response. Gene sets typically upregulated during osmotic shock and oxidative stress exposure are also upregulated in the *ΔΔtlo* mutant, which typically would not be upregulated in a healthy cell in YEPD.

Another gene set that is upregulated in the *ΔΔtlo* mutant compared to the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains is the "MED31_DN" gene set, which includes genes that were found to be downregulated in a *ΔΔmed31* mutant *C. albicans* strain. Both the transcriptomes and the phenotypes of the *ΔΔtlo* mutant and the *ΔΔmed31* mutant are similar. Med31 is another mediator subunit, part of the middle domain, and the *ΔΔmed31* mutant was defective for filamentation, biofilm formation and was more susceptible to cell wall damaging agent, Congo Red (Uwamahoro *et al.*, 2012). These phenotypic deficits are also evident in the *ΔΔtlo* mutant, and are restored upon expression of *TLOα1* or *TLOβ2*, as shown in Chapter 5. The *ΔΔmed31* mutant strain showed reduced expression of the *TYE7* transcription factor (Uwamahoro *et al.*, 2012), which was also seen in the *ΔΔtlo* mutant. From the microarray data published on the *ΔΔmed31* mutant, there does not appear to be any difference in expression of the *TLO* gene family itself, or any of the other Mediator subunits, however, this differential expression in this strain was compared to a *ΔΔmed31+MED31* complemented strain and not the WT parent (Uwamahoro *et al.*, 2012). There does not appear to be any data on the structure of the Mediator complex in the *ΔΔmed31* mutant, specifically if the Med2 homologues, the Tlos, can incorporate into a mediator complex lacking Med31. However, these data highlight that the deletion of the *TLO* gene family disrupts the function of the Mediator complex.

The morphology of the strains expressing *TLOβ2* in the *ΔΔtlo* mutant was an interesting phenotype that came to light upon examination of the strains generated in this work. In this RNA-seq analysis, there were somewhat mixed results returned with regards to hyphal morphology. The GSEA results identified that some genes upregulated in hyphae were enriched in the *ΔΔtlo::P_{TET}TLOβ2* strain, however some also identified genes downregulated in hyphae to be highly expressed in this strain. This could be due to only a small percentage of the cells in *ΔΔtlo::P_{TET}TLOβ2* being truly hyphal, for example, around 6% were true hyphae in an overnight culture grown in YEPD at 37 °C, see Chapter 5. The *P_{ENO}TLOβ2* strain could be a better candidate for investigating this phenotype as it appeared to be more hyphal than the *ΔΔtlo::P_{TET}TLOβ2* strain, around 20% of cells were true hyphae in an overnight culture grown at 37 °C in YEPD. The induction of hyphal growth in the *TLOβ2* reintroduction strains was more efficient than AHY940, so perhaps an RNA-seq experiment looking at transcriptional changes during the early stages of hyphal induction may help us understand the role that *TLOβ2* plays in the formation of hyphae. In the *ΔΔtlo::P_{TET}TLOα1* strain, there was enrichment of gene sets "HYPHAE_FBS_37_DN" and "HYPHAE FBS37 (NANTEL)_DN" compared to the *ΔΔtlo*

mutant, and these sets were more significantly enriched than in the *ΔΔtlo::P_{TET}TLOβ2* strain compared *ΔΔtlo* mutant. The morphologies of these strains are all somewhat different, the *ΔΔtlo* mutant grows pseudohyphally, the *ΔΔtlo::P_{TET}TLOα1* grows as typical yeast blastospores, and the *ΔΔtlo::P_{TET}TLOβ2* strain grows as blastospores with a subset of cells growing as hyphae. The increased expression of these gene sets containing gens that are typically downregulated in hyphae point toward a genetic basis for the blastospore morphology in the *ΔΔtlo::P_{TET}TLOα1* and in the majority of cells in the *ΔΔtlo::P_{TET}TLOβ2* strain.

There was also significant, unique, enrichment of the "UME6_BIND" gene set in the *ΔΔtlo::P_{TET}TLOβ2* strain. This gene set contains genes found to interact with Ume6 in a ChIP-chip microarray performed on this protein in *S. cerevisiae*. In *C. albicans*, the *UME6* gene is known to be an important regulator in the hyphal morphogenesis pathway, and while a *ΔΔume6* mutant can still form germ tubes in the early stages of hyphal induction, the hyphae cannot elongate (Banerjee *et al.*, 2008). Experiments investigating the effects of overexpressing *UME6* found that increasing the expression of this transcriptional regulator increases hyphal formation and increases the virulence of strain in a murine infection model (Carlisle *et al.*, 2009). The unique enrichment of genes bound by this regulator could indicate a role for *TLOβ2* in this part of the hyphal morphogenesis pathway, perhaps a direct interaction between Tloβ2 and Ume6, and give some explanation for the formation of hyphae in this strain in the absence of inducing conditions.

The amount of gene sets found to be enriched in both the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains compared to the *ΔΔtlo* mutant indicates that there is some level of shared functionality between these *TLOs*. However, the large number of unique gene sets enriched in the *ΔΔtlo::P_{TET}TLOα1*, not enriched in the *ΔΔtlo::P_{TET}TLOβ2* strain, compared to the *ΔΔtlo* mutant suggests that there are specific functions carried out by only *TLOα1* that cannot be carried out by *TLOβ2.* The *ΔΔtlo::P_{TET}TLOβ2* strain itself also has its own group of uniquely enriched gene sets compared to the *ΔΔtlo* mutant that were not enriched in the other strains, this too indicates specialised roles for this *TLO* within the cell.

While phenotypic and transcriptomic analyses suggest significant roles for *TLOα1* and *TLOβ2*, the function of the *TLOγ11* gene remains somewhat of a mystery. With no discernible effects found during phenotypic testing, the RNA-sequencing reflected this,

with only comparatively small differences in expression observed. Gene sets that were found to be enriched in the *ΔΔtlo::P_{TET}TLOγ11* strain compared to the *ΔΔtlo* mutant were those containing genes upregulated upon exposure to xenobiotic stress and heat shock, as well as significant enrichment of genes related to the proteasome. There was enrichment of the "MED31_UP" gene set in the *ΔΔtlo::P_{TET}TLOγ11* strain, but while this enrichment had a p value of <0.05 (0.009), the false discovery rate was high, leading to a family wise error rate p value of 1, meaning that this is not a reliable result and possibly a false discovery. This indicates that Mediator function is not restored in the *ΔΔtlo::P_{TET}TLOγ11* strain, while expression of *TLOα1* and *TLOβ2* in the *ΔΔtlo* background does restore Mediator function, bringing into to question the ability of Tloγ11 to incorporate into the complex in this strain. It should be noted that the level of Tloγ11 protein present in this strain has been questioned previously. In contrast the sets upregulated in the *ΔΔtlo* mutant included the "RHE90_UP" set and many sets related to carbohydrate transport. The enrichment of the group of genes related to the proteasome (GO:0000502) could indicate an increased level of protein degradation occurring in the *ΔΔtlo::P_{TET}TLOγ11* strain. The proteasome of *C. albicans* is a multiprotein complex made up of a 20S core and two 19S regulatory subunits. Proteins are targeted for degradation by the proteasome by ubiquitinoylation, detected by the regulatory subunits, and then hydrolysed by the core (Hossain *et al.*, 2020). In *C. dubliniensis* it is not possible to build up a free pool of CdTlo1 protein in the cells without the use of a proteasome inhibitor. Only CdTlo1 protein which is associated with Mediator is protected from degradation. Otherwise this protein is subjected to rapid proteasome-dependent degradation. This is not the case for the CdTlo2 protein, and a free pool of this protein can be built up by over expression of the gene without the need for treatment of the cells with proteasome inhibitor. This indicates that Tlo proteins in *Candida* can be subject to sequence specific proteasome-dependent degradation (Liu *et al.*, 2016). The possibility of increased protein degradation within the *ΔΔtlo::P_{TET}TLOγ11* cells could indicate why the Tloγ11 had been hard to detect by Western Blot as described in earlier chapters. This effect also appears to be sequence specific, where the Tloα1 and Tloβ2 proteins are not subject to this degradation. It would be of interest to treat *ΔΔtlo::P_{TET}TLOγ11* cells with a protease inhibitor prior to protein extraction to determine if the Tloγ11 protein could then be detectable. In *C. dubliniensis* the cell tightly controls the level of CdTlo present, and ensures there is no free pool of CdTlo1 protein. Treatment with a protease inhibitor can artificially generate a free pool of this protein, indicating that degradation is proteasome dependent (Liu *et al.*, 2016). It could be the case that the cell requires some method of

post-transcriptionally controlling the amount of Tlo protein in the cell, and activation of the proteasome could be one method of doing this.

It would also be interesting to see if the expression of another *TLO* with *TLOγ11* in the *ΔΔtlo* background would repress these proteasome genes, or whether the expression of *TLOγ11* would still induce the transcription of these genes and affect the second Tlo protein.

6.4.3 Transcriptomic response of AHY940, ΔΔtlo *mutant and* TLO *reintroduction strains to oxidative stress induced by tBOOH*

The transcriptomic response to oxidative stress was examined in the AHY940 strain, the *ΔΔtlo* mutant and the *TLO* reintroductions strains by sequencing the transcriptome after 30 minutes of exposure to tBOOH, and subsequently comparing gene expression to that in YEPD. tBOOH is an oxidative stress generating agent and it is thought that exposing cells to tBOOH generates a similar response as if the cells were phagocytosed by macrophages. The ROS generated damage lipids in the cell walls by peroxidation (Fekete *et al.*, 2007).

Comparison of the *ΔΔtlo::*P_{TET}*TLOγ11* strain in these conditions was not successful and no genes in the differential expression analysis were significantly differentially expressed. This may be down to a lack of reproducibility between isolates, as indicated in the PCA analysis performed by Novogene, where both the YEPD samples and the tBOOH samples are less closely clustered than all other strains.

In all strains, apart from *ΔΔtlo::*P_{TET}*TLOγ11,* GSEA highlighted that the "XS_UP" gene set was enriched in tBOOH conditions. This gene set contains genes found to be upregulated in *C. albicans* when exposed to 0.4 mM H_2O_2 in YEPD at 30 °C, which are slightly different conditions to those used in the experiments here (Enjalbert *et al.*, 2003), although finding the "XS_UP" gene set enriched shows that the tBOOH is induced a very similar oxidative stress response as H_2O_2.

Examining the expression change in the genes in the "XS_UP" gene set confirmed that the majority of these genes were upregulated in the tBOOH conditions compared to in YEPD in all strains except *ΔΔtlo::*P_{TET}*TLOγ11.*The mean expression of these genes was higher in tBOOH than in YEPD. Hierarchical clustering performed on the expression sets

grouped the responses of the *ΔΔtlo* mutant and the *ΔΔtlo::P_{TET}TLOα1*strain grouped together, as did the AHY940 and *ΔΔtlo::P_{TET}TLOβ2* strains. This is not congruent with the phenotype, where AHY940 and the *ΔΔtlo::P_{TET}TLOα1*strain had similar high tolerance levels. The *ΔΔtlo* mutant and the *ΔΔtlo::P_{TET}TLOγ11* had similar low tolerance levels, and the *ΔΔtlo::P_{TET}TLOβ2* displayed an intermediate tolerance level to oxidative stress induced by tBOOH. The expression of genes within the "XS_DN" set were also examined in the strains under oxidative stress. It was seen that in the *ΔΔtlo* mutant, the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* that these genes were generally downregulated in tBOOH, as expected. However, there was a mean increase in expression of these genes in the AHY940 strain, which was not expected. Upon further inspection, it was found that about 10% of these genes in this category are located on chromosome 5, which has been shown to be present in three copies in the AHY940 strain, which could be influencing the expression of these genes. When the expression levels of the gene that fall under the "response to oxidative stress" GO term (GO:0006979) were analysed. It was seen that these genes were generally upregulated in all strains (except *ΔΔtlo::P_{TET}TLOγ11*), with hierarchical clustering again revealing that the *ΔΔtlo* mutant and the *ΔΔtlo::P_{TET}TLOα1*strain responses grouped together, as did those of AHY940 and *ΔΔtlo::P_{TET}TLOβ2.*

The fact that the *ΔΔtlo* mutant elicits a similar response to oxidative stress as the AHY940, *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains, but that it is much more sensitive than these strains is interesting. This indicates that the defects in the *ΔΔtlo* mutant could be the cause of the increased sensitivity, or it could be a specific part of the oxidative stress response that is not being induced in the *ΔΔtlo* mutant. Studies suggest that the opaque cell phenotype is more resistant to oxidative stress than the white type and that because of this the opaque cell is more resistant to phagocytosis by macrophages (Alby and Bennett, 2009). It was previously speculated that the *ΔΔtlo* mutant existed predominantly as opaque cells, see Section 6.4.2. If this was the case it would mean that the *ΔΔtlo* mutant would be more resistant than the other strains because of its opaque morphology. This is not the case. However, it was found that there was an increase in expression of opaque specific genes in the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains upon exposure to tBOOH. This was not the case in AHY940. Oxidative stress has been shown to increase phenotypic switching in *C. albicans* (Alby and Bennett, 2009). The increase in opaque specific genes was not found in the *ΔΔtlo* mutant, but this may be due to the already high level of expression of these genes in this strain in YEPD.

The *ΔΔtlo* mutant was previously shown to be defective in *ERG11* expression. This could be partly responsible for the increased susceptibility of this strain to oxidative stress. *ERG11* has been found to be coupled with the oxidative stress response in *C. albicans*, with *ΔΔerg11* mutants being more sensitive to oxidative stress than WT strains (Wu *et al.*, 2018). The fact that *ERG11* expression is restored in the *ΔΔtlo::*P_{TET}*TLOα1* and *ΔΔtlo::*P_{TET}*TLOβ2* strains and not in the *ΔΔtlo::*P_{TET}*TLOγ11* strain correlates with the abilities of these strains to tolerate oxidative stress, where *ΔΔtlo::*P_{TET}*TLOα1* and *ΔΔtlo::*P_{TET}*TLOβ2* are more tolerant than the *ΔΔtlo* mutant, and the *ΔΔtlo::*P_{TET}*TLOγ11* strain does not display a difference in tolerance from the *ΔΔtlo* mutant.

Expression of the *SSU81* gene, also known as *SHO1*, was found to be downregulated in tBOOH in the *ΔΔtlo* mutant. This gene encodes a protein that mediates resistance to oxidative stress through activation of the Cek1 MAP kinase. This gene was upregulated upon challenge with oxidative stress in AHY940, *ΔΔtlo::*P_{TET}*TLOα1* and *ΔΔtlo::*P_{TET}*TLOβ2*. *SSU81/SHO1* mutants cannot filament under certain conditions (e.g. in Spider) and are more susceptible to Congo Red and Calcofluor White (Roman *et al.*, 2005), which again were phenotypes found in the *ΔΔtlo* mutant compared to AHY940, *ΔΔtlo::*P_{TET}*TLOα1* and *ΔΔtlo::*P_{TET}*TLOβ2*. It is possible that the *TLO* gene family could play a role in activating this gene under oxidative stress conditions.

These differences and accumulation of defects, such as reduced *ERG11* and *SSU81/SHO1* expression, could make all the difference in the reduced ability of the *ΔΔtlo* mutant strain to resist oxidative stress. Even though the *ΔΔtlo* mutant responds to exposure to tBOOH by activating many of the genes that are typically activated during exposure to oxidative stress inducing compounds, this is clearly not sufficient to permit the same level of growth in the presence of tBOOH as the other strains tested. It has already been shown that the *ΔΔtlo* mutant is severely defective in many phenotypes tested, and so it is not surprising that the strain is much more sensitive to oxidative stress than the others. It would be interesting to artificially overexpress genes such as, *ERG11* or *SSU81/SHO1*, in the *ΔΔtlo* mutant to determine if this could restore the ability to tolerate oxidative stress.

In YEPD alone the *ΔΔtlo* mutant exhibits a transcriptional response similar to a cell undergoing oxidative stress and nutritional starvation, as evidenced by the upregulation of gene sets that are induced during exposure to oxidative stress and genes

that are upregulated upon engulfment by host immune cells. The *ΔΔtlo* mutant is already an unfit cell with an irregular transcriptome, and while it is undergoing a transcriptional response when exposed to tBOOH, it appears that this is not enough to rescue the cell and tolerate the stress.

The intermediate level of oxidative stress tolerance displayed by the *ΔΔtlo::P_{TET}TLOβ2* was also an interesting phenotype, and the transcriptomic response of this strain was similar to that of AHY940 and *ΔΔtlo::P_{TET}TLOα1*. From data generated during phenotypic testing of the strains, it appears that the *ΔΔtlo::P_{TET}TLOβ2* strain (and to a greater extent the *ΔΔtlo::P_{ENO}TLOβ2* strain) form hyphae in non-inducing conditions, and they appear to more readily form hyphae under inducing conditions. The possibility that the hyphal morphology of the strains themselves is causing increase susceptibility to tBOOH has been discussed previously, Chapter 5. When *C. albicans* strains were exposed to subinhibitory levels of tBOOH and selected based on resistance, it was found that strains that have evolved to be more tolerant of the stress were less able to form hyphae under inducing conditions (Fekete *et al.*, 2007). On top of this there is also the possibility that exposure to tBOOH also promotes the switch to the hyphal morphology. Filamentation is induced upon phagocytosis of *C. albicans* cells by immune cells and acts as an escape mechanism from oxidative killing by these immune cells, puncturing the cellular membrane, and killing the immune cell (Dantas *et al.*, 2015, Lorenz *et al.*, 2004). In the RNA-seq data, the *HSP12* gene was shown to be highly upregulated in the *ΔΔtlo::P_{TET}TLOβ2* under oxidative stress conditions compared to in YEPD, and this level of increase is only seen in this strain. There is no significant difference in its expression in AHY940 or the *ΔΔtlo* mutant, there is an increase in *HSP12* expression in the *ΔΔtlo::P_{TET}TLOα1* strain, however the increase is not as dramatic as was seen in the *ΔΔtlo::P_{TET}TLOβ2* strain (+2.2 L_2FC in *ΔΔtlo::P_{TET}TLOβ2* vs. +0.9 in *ΔΔtlo::P_{TET}TLOα1*). This gene is typically upregulated in response to oxidative stress and has been shown to be regulated by the Efg1 transcription factor, and while the gene itself is not essential for hyphal formation, its overexpression has been shown to enhance *C. albicans* filamentation in the early stages (Fu *et al.*, 2012).

Comparison of the transcriptomes of the strains exposed to tBOOH to each other highlighted that the *ΔΔtlo* mutant is expressing genes typically upregulated upon exposure to oxidative stress, while comparatively AHY940 is expressing these genes at the lowest level (along with the dubious *ΔΔtlo::P_{TET}TLOγ11* strain). The *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains appear to express these genes at an intermediate level

compared to AHY940 and the *ΔΔtlo* mutant. This could indicate that the stress induced by tBOOH exposure for 30 minutes is not eliciting a strong response in the WT, but the exposure is affecting the *ΔΔtlo* mutant dramatically. This could be due to the fact that the *ΔΔtlo* mutant is already expressing these genes related to oxidative stress in YEPD with no stress inducing agent present. The upregulation of genes related to the oxidative stress response were also found to be upregulated in the *Δtlo1/Δtlo2 C. dubliniensis* mutant strain compared to the *C. dubliniensis* WT (Haran *et al.*, 2014). This could mean that there is already a high level of ROS within these *ΔΔtlo* mutant cells, which is restored to normal levels by the reintroduction of *TLOα1* and *TLOβ2*, and because of this the *ΔΔtlo* mutant cannot tolerate additional stress from tBOOH, despite the upregulation of oxidative stress response genes. Differences in baseline levels of ROS in the *ΔΔtlo::*P_{TET}*TLOα1* and *ΔΔtlo::*P_{TET}*TLOβ2* strains could be the reason for the different levels of tolerance to tBOOH where there does not appear to be major transcriptional differences. Investigation into the levels of ROS within the cells in YEPD and upon exposure to tBOOH, which could be performed with dyes to measure ROS levels within cells (James *et al.*, 2015), could explain the differences in tolerance seen.

6.4.5 Closing Remarks

Transcriptomic analysis described in this chapter was performed to attempt to understand the role of the *TLO* gene family in *C. albicans* using the *TLO* mutant strains generated as described in previous chapters. It was clear from phenotypic testing performed on the *ΔΔtlo* mutant that this strain was defective in most normal cellular processes and in the ability to tolerate oxidative stress or cell wall perturbation. The transcriptomic analysis of this strain showed that this strain seems to exist in a state of perpetual stress. The transcriptome very much resembled the "starvation transcriptome" of the *Δtlo1/Δtlo2 C. dubliniensis* mutant described by Haran *et al.* in 2014, as well as the phenotypes of these two strains being almost identical (Haran *et al.*, 2014). Reintroduction of the *TLOα1* gene in to the *ΔΔtlo* background was able to restore all phenotypes tested, however the transcriptome of this strain was still different to that of the AHY940 parent. It is possible that the chromosome 5 trisomy is influencing this analysis, so a repeat experiment using a newly generated *Δleu2/LEU2* heterozygous derivative of SC5314 may clear up any ambiguity here. Although some potential reasons to explain the morphological phenotype of the *ΔΔtlo::*P_{TET}*TLOβ2* have been outlined above, a clearer picture could emerge by sequencing the transcriptomes of the more dramatically hyphal *ΔΔtlo::*P_{ENO}*TLOβ2* strain. Another possibility could be selectively

sequencing only the hyphal cells in this strain, which would require some sort of filtering to retain only the hyphal cells, perhaps based on size. This could uncover some expression patterns that could be leading to the hyphal morphology of the cells being missed in this analysis where most of the cells are in the yeast morphology. Another experiment that could help uncover a role for *TLOβ2* in the hyphal morphogenesis pathway would be to sequence the RNA of the strains at different time points after induction of the hyphal switch, such as induction with temperature increase and FCS, or by growing cells in Spider medium. The fact that *TLOγ11* did appear to be altering the expression of some genes when introduced into the *ΔΔtlo* background, albeit a very small number of genes, indicates that this gene is carrying out some functions within the cell, but that these effects are not visible on a phenotypic level.

Analysis of the transcriptomic responses of the strains to oxidative stress exposure was also performed. While definitive analysis of *ΔΔtlo::*P_{TET}*TLOγ11* strain was not possible, some interesting findings about the other strains came to light. The *ΔΔtlo* mutant and the *ΔΔtlo::*P_{TET}*TLOβ2* strain having very similar transcriptomic responses as the AHY940 WT and the *ΔΔtlo::*P_{TET}*TLOα1,* but having very different tolerance levels to tBOOH points to the *TLO* gene family not having a direct role in affecting transcription in the oxidative stress response pathways, but more so having an effect on other features of the cell to such a degree that even the correct induction of the oxidative stress response cannot tolerate the stress. Findings from this analysis would point towards defects in cell wall and cell membrane biogenesis, specifically sterol and lipid biogenesis as being one of the main reasons why the *ΔΔtlo* mutant, and the *ΔΔtlo::*P_{TET}*TLOγ11* and *ΔΔtlo::*P_{ENO}*TLOγ11* strains, are much more sensitive to oxidative stress. The cell wall and cell membrane composition of the *ΔΔtlo::*P_{TET}*TLOβ2* strain may also be composed differently such that the cells are so much more sensitive to oxidative stress induced by tBOOH than AHY940 and *ΔΔtlo*::P_{TET}TLOα1. Further investigation into the cell wall and cell membrane contents of the strains could bring to light any compositional differences that may be leading to differences in stress tolerance, and perhaps explain other phenotypes, such as the increased sensitivity to cell wall perturbing compounds and the azole resistance phenotypes.

Chapter 7

Mapping Tlo interactions and RNA Polymerase II interactions with the genome in *TLO* mutant strains via ChIP-seq

7.1 Introduction

Chromatin Immunoprecipitation (ChIP) allows for the study of interactions between proteins and DNA. By crosslinking cells with formaldehyde, preserving protein:DNA interactions, and using a specific antibody to pull down a protein of interest, the DNA bound by this protein can be isolated. By pairing this technique with DNA sequencing, sites where a protein interacts with the DNA can be identified. This process can be used to map interactions of transcription factors to look at their regulatory functions, and can also be used to investigate histone modifications, their locations, and their epigenetic consequences (Schmidt *et al.*, 2009). Similar to RNA-seq analysis, a precursor to ChIP-seq was ChIP followed by analysis on a microarray, also termed ChIP-chip. But again ChIP-seq is a much more sensitive approach and can detect more binding peaks with much higher resolution (Ho *et al.*, 2011).

In *Candida* spp., ChIP-chip has been used widely to map protein:DNA interactions. Some examples include mapping the interaction sites of the Snf6 protein by Tebbji *et al.* in 2017, a subunit of the fungal SWI/SNF complex involved in chromatin remodelling (Tebbji *et al.*, 2017), and analysing the binding locations of the Tac1 transcription factor by Liu *et al.* in 2007 to determine involvement in azole resistance (Liu *et al.*, 2007). In a study by Nobile *et al.* in 2012, attempts to decipher the transcriptional network that regulates biofilm formation in *C. albicans* was investigated by performing ChIP-chip analysis using each of their candidate regulatory proteins as the protein of interest to map their interaction sites, further validating the transcriptional network they uncovered (Nobile *et al.*, 2012).

The RNA Polymerase II (RNAP) complex is made up of 12 different subunits which together act to transcribe all protein coding RNAs in the cell, as well as many non-coding RNAs. This enzyme acts in congruence with the preinitiation complex, made up of many other factors, including the Mediator, to transcribe genes (Schier and Taatjes, 2020). It has been proposed that ChIP-seq targeting the RNAP complex could act as a cheaper alternative to RNA-seq. If adequate sensitivity can be achieved, RNA Polymerase could be mapped to all genes being actively transcribed, and the expression level quantified by measuring RNAP occupancy at each gene (Tan and Wong, 2019). In *C. albicans,* many studies have used ChIP-seq to screen mutants for differences in RNAP occupancy and transcriptional activity. Cas5 is a transcriptional regulator important in the response to cell wall stress, and in a *ΔΔcas5* mutant RNAP occupancy was dramatically different

from that in WT. Analysis of the functions of the genes occupied and unoccupied differentially in the strain allowed for conclusions to be made about the role of Cas5, for instance that it is a regulator of cell cycle dynamics under stress conditions and basal conditions (Xie *et al.*, 2017).

RNA polymerase II ChIP-seq was used in the current study, initially as a control to validate development of the protocol, however, this was expanded into its own experiment to investigate the differences in RNA Polymerase II occupancy in the *ΔΔtlo* mutant and the AHY940 parent.

ChIP-chip analysis performed by Haran *et al.* in 2014 showed that in *C. dubliniensis* the CdTlo1 protein localises at repeat sequences such as the telomeres and MRS (when grown in YEPD at 30 °C). They also mapped CdTlo1 binding to ORFs. These ORFs had varying levels of expression, with some genes being repressed in YEPD, and some being actively transcribed, indicating a dual role for CdTlo1 in activation and repression. Genes involved in glycolysis and amino acid metabolism were found to be bound by CdTlo1, and these genes were not expressed in a *Δtlo1/Δtlo2* mutant *C. dubliniensis*, suggesting a direct role for CdTlo1 in the expression of these genes. Genes related to the starvation response and nitrogen scavenging, which were upregulated in the *Δtlo1/Δtlo2* mutant, were also found to be bound by CdTlo1, indicating that the presence of CdTlo1 is essential for their silencing (Haran *et al.*, 2014).

7.1.1 Aims of this work

The main aims of this chapter were to use ChIP-Seq based approaches to further investigate the role of Tlos in *C. albicans* using strains described earlier in this work (i.e. the *ΔΔtlo* null mutant, the strains where representative *TLOs* were reintroduced under the control of the *TET1* promoter, and the AHY940 WT parent). An anti-HA antibody was used to map Tlo interactions with the genomes of the relative strains via the HA tag introduced to the proteins upon reintroduction into the *ΔΔtlo* background. In addition, an anti-RNA Polymerase II ChIP-seq was performed on the *ΔΔtlo* strain and the AHY940 strain to investigate the activity of this complex in *C. albicans* with and without *TLO*s. The data gathered in this chapter was also compared to the transcriptomic data collected in Chapter 6 to determine if there was a correlation between protein interaction and transcription in the strains.

7.2 Materials and Methods

7.2.1 Growing and fixing cells

The protocol for growing, fixing and spheroplasting *C. albicans* cells was adapted from the method described by Haran *et al.* for ChIP-chip analysis in *C. dubliniensis* (Haran *et al.*, 2014).

A single *Candida albicans* colony was picked from a YEPD plate (24 – 48h old) and grown overnight in 4 ml cultures of YEPD in 13 ml tubes at 37 °C in a 200 rpm shaking incubator. The concentration of cells was quantified by spectrophotometry, the cells were added to 25 ml YEPD broth in an Erlenmeyer flask to a density of OD_{600} = 0.2. This was returned to a 37 °C shaking incubator at 200 rpm until the density reached OD_{600} = 2.0. Cells were transferred to a 50 ml tube and pelleted at 2,000 x *g* for 5 minutes, the supernatant was discarded. Then the cells were washed with 5 ml PBS and spun down as before, and supernatant discarded. Cells were then resuspended in 36 ml PBS with the addition of 1 ml 37% formaldehyde for crosslinking. This was incubated in a 200 rpm shaking incubator at 37 °C for 30 min. The crosslinking reaction was then quenched with 2 ml of 2.5 M glycine and returned to the shaking incubator for 10 min. The cells were then pelleted as before, and the supernatant was discarded. These pelleted cells could then be stored at 4 °C for up to 48 hours before spheroplasting.

7.2.2 Spheroplasting

Pelleted, fixed cells were resuspended in 3 ml Spheroplasting Buffer 1 (10 mM DTT, 200 mM Tris [pH 9.5]) and incubated at 37 °C for 15 min in a 200 rpm shaking incubator. Cells were pelleted and resuspended in 5 ml Spheroplasting Buffer 2 (1.2 M Sorbitol, 20 mM HEPES [pH 7.4]). Cells were pelleted at 2,000 x *g* for 5 min, the supernatant was discarded, and the cells were resuspended again in 5 ml Spheroplasting Buffer 2, with the addition of 0.4 mg/ml Zymolyase T100 and 0.5M 4-(2-aminoethyl-benzenesulfonyl fluoride (AEBSF). Cell suspensions were transferred to a flat-bottomed Erlenmeyer flask and incubated in a 37 °C shaking incubator for a maximum of 2.5 h or until >50% of cells were deemed to be spheroplasts (i.e. they lost the rigidity of their cell wall and appeared spherical compared to the more oval shaped yeast cells).

After spheroplasting, 5 ml of pre-chilled Post Spheroplasting Buffer (1.2 M Sorbitol, 1 mM $MgCl_2$, 20 mM PIPES [pH 6.8]) was added. The samples were transferred to a 50 ml tube and spun down at 2,000 x *g* for 5 min. The supernatant was discarded, and the cells were washed with 5 ml PBS. Suspensions were spun down at 2,000 x *g* for 5 min and supernatant was discarded. These pellets were held on ice until proceeding to the MNase digestion.

7.2.3 MNase digestion

ChIP-seq requires that the DNA be sheared to small sizes to map protein interactions with greater resolution. The protocol described by Haran *et al.* used sonication for digestion of DNA (Haran *et al.*, 2014), however here digestion was performed using mung bean micrococcal nuclease to fragment the DNA reliably and reproducibly. The MNase digestion protocol used in these experiments was modified from the method described by Kent *et al.* in 2011 (Kent *et al.*, 2011).

Cell spheroplast pellets were resuspended in 950 µl of pre-chilled 1 M Sorbitol and spun down at 2,000 x *g* for 5 min and the supernatant discarded. Cells were kept on ice as much as possible. The samples were then resuspended in 600 µl of Spheroplast Digestion Buffer (1M Sorbitol, 50 mM NaCl, 10 mM Tris-HCl [pH 7.5], 5 mM $MgCl_2$, 1 mM $CaCl_2$, 1 mM β-mercaptoethanol, 0.5 mM Spermidine, 0.075% Tergitol) and 200 µl aliquots of this suspension were incubated with various concentrations of MNase ranging from 1 to 10 U (10 U/ul stock in MNase buffer [10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 100 µg/ml bovine serum albumin]) in 1.5 ml Eppendorf tubes. The samples were incubated at 37 °C for 1 – 5 min. Examination of electrophoresed samples allowed the visualisation of the digestion products and it was determined that the optimum conditions were 7 U of MNase for 4 min. After the desired digestion time had elapsed, the digestion was terminated by the addition of 20 µl 250 mM EDTA to each sample. Samples could be stored at -20 °C or could be kept on ice before samples checks.

7.2.4 Reverse crosslinking, Proteinase K treatment and DNA purification for sample checks

Adequate sample digestion was optimised and confirmed by purifying DNA from the MNase digestions, with reverse crosslinking and Proteinase K treatment steps. This

protocol was adapted from (Wal and Pugh, 2012). Briefly, 3 µl of Proteinase K (20 mg/ml stock) was added to the samples and incubated at 65 °C overnight. DNA was extracted with 450 µl PCIA, vortexed for 20 seconds and centrifuged at 14,000 rpm for 6 min. The upper aqueous layer was removed to a new Eppendorf tube where 1 µl of glycogen (20 mg/ml) and 0.6 volume of isopropanol was added. Samples were mixed well by inversion and incubated at room temperature for 30 min. Samples were then centrifuged at 14,000 rpm for 30 min at room temperature. Pellets were washed in 500 µl of room temperature 70% ethanol and centrifuged for 10 min at 14,000 rpm at room temperature. Supernatant was discarded and the pellets were allowed to dry for 10 – 15 min. Pellets were resuspended in water with 50 µg/ml RNase (10 mg/ml stock) and incubated at 37 °C for up to 2 h, but for no less than 30 min. DNA concentrations were quantified using a Nanodrop 2000 and bands were visualised using agarose gel electrophoresis, see Section 2.3.4. Optimal digestion was classified as the absence of high MW bands, and the presence of the most intense bands at approximately 147 bp.

7.2.5 Immunoprecipitation

This immunoprecipitation protocol used was adapted from the method described in Haran *et al.* 2014 (Haran *et al.*, 2014).

Stored lysates were thawed on ice and made up to a volume of 500 µl with FA Lysis Buffer (50 mM HEPES [pH 7.5], 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Sodium Deoxycholate) with 1X protease inhibitor. 20 µl of this was removed to a separate tube labelled “input”. This sample was used as a control and did not undergo immunoprecipitation. To the remaining sample, 4.5 µl of antibody (Anti-RNA Polymerase II, CTD, cl8WG16 or Anti-HA 12CA5) was added and incubated overnight on a rotating wheel at 20 rpm at 4 °C. The input sample was treated with pronase.

To the 20 µl of input sample the following were added: 100µl of ChIP elution buffer (50 mM Tris-Cl [pH 7.5], 10 mM EDTA, 1% SDS), 60 µl of TE buffer (pH 7.5) (10 mM Tris [pH 7.5], 1 mM EDTA), 20 µl Pronase (from 20 mg/ml stock) and 1 µl 1M $CaCl_2$. The sample was then processed on a thermocycler for 2 h at 42 °C followed by 6 h at 65 °C. Samples were held overnight at 10 °C and then stored at -20 °C until DNA purification.

After overnight incubation of the lysate with the antibody, immunoprecipitation was performed. A master mix of Protein A and Protein G Dynabeads (Invitrogen), 20 µl of

each Dynabead solution per sample, was made up in an Eppendorf tube and 1 ml of FA Lysis Buffer with 1X PI mix was added. A small portion of each pipette tip was cut off when working with the beads. This master mix was incubated on a rotating wheel at 20 rpm at room temperature for 5 min. Beads were captured with a magnetic rack (New England Biolabs) and the supernatant was removed. The amount of time that beads were outside of solution was minimised to prevent drying out. The beads were washed in 1 ml of FA Lysis buffer with 1X PI twice more as above and then finally resuspended in 1 ml of FA Lysis buffer with 1X PI. The bead solution was evenly distributed between the lysates, and these were incubated on a rotating wheel at 20 rpm for 2 h at 4 °C.

The beads were then captured on a magnetic rack and subjected to several washing steps. For each wash 1ml of buffer was added and the sample incubated on a rotating wheel at 20 rpm for 5 min at room temperature, after which the beads were captured again, and the supernatant discarded. Wash steps were as follows: 1 wash in FA Lysis buffer with no PI, 2 washes in Wash Buffer 1 (FA Lysis buffer with 0.5 M NaCl), 2 washes in Wash Buffer 2 (10 mM Tris-Cl [pH 7.5], 0.25 M LiCl, 1 mM EDTA, 0.5% Tergitol, 0.5% Sodium Deoxycholate) and finally 1 wash in TE buffer (pH 7.5).

After the final wash step the protein:DNA complexes were eluted from the beads. 250 ul of ChIP elution buffer was added and the samples were vortexed vigorously. Samples were incubated at 65 °C for 20 min, then incubated on a rotating wheel at 20 rpm for 10 min at room temperature. The samples were then centrifuged at 13,000 rpm for 2 min and the supernatant (around 250 µl) was collected. This is the immunoprecipitate (IP).

The IP was then pronase-treated. 50 µl of Pronase (20 mg/ml stock) was added to the IP, along with 1.5 µl of 1 M $CaCl_2$. This was processed on a thermocycler using the same settings used for the input sample, above. The pronase treated IP samples were then stored at -20 °C until DNA purification.

7.2.6 DNA purification and sequencing

DNA purification was performed using a QIAquick PCR Purification Kit (QIAGEN) according to the manufacturer's instructions. Samples were quantified using a Qubit (Thermo Fisher Scientific). Samples (10 ng) were sent to Novogene for DNA sequencing on the Illumina platform. Two samples were sent to Genewiz for sequencing as a trial, a

single RNAP ChIP from each of AHY940 and *ΔΔtlo,* however all other samples were sequence by Novogene.

7.2.7 ChIP-seq data analysis

Commands for each step of the analysis can be found in Appendix X.

Illumina paired end sequencing was performed on the samples by Novogene to a depth of about 40 to 50 million raw reads per sample. For the two test samples sent to Genewiz, a sequencing depth of 7 to 8 million raw reads were outputted. Some quality checks were performed by the sequencing company before data release to ensure samples were sequenced to the desired depth and the quality was acceptable. Fastq files were also run through FastQC software (available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to confirm quality. The fastp tool (Chen *et al.*, 2018) was used for quality filtering and Illumina adaptor trimming.

For alignment both file pairs for each sample (from paired-end sequencing) were aligned to the reference genome (as used in Chapter 4) in their zipped format (fq.gz) using the Bowtie2 tool (Langmead and Salzberg, 2012), followed by Samtools (Danecek *et al.*, 2021) for sorting and indexing.

MACS2 was used to call peaks in the IP samples relative to the control input samples (Zhang *et al.*, 2008). Broad peaks were called for the RNAP samples, and narrow peaks for the anti-HA samples. Broad peaks detect regions of enrichment, while narrow peaks detect more specific peaks of enrichment, typically used to analyse transcription factor binding.

Replicates were analysed for consistency using bedtools intersect, part of the bedtools suite (Quinlan and Hall, 2010). The MACS2 tools was then rerun to combine the replicate bam files and output a single peak file. The peak files (which are in BED6+4 format) were converted to BED6 format for viewing in the Integrated Genome Browser (Freese *et al.*, 2016) using the bedtools suite.

Bedtools intersect was used to determine if peaks from the sample files intersected the TSS of genes, the promoter regions or the ORF of genes in the reference genome. These

lists of genes were then subjected to further analysis, including GO analysis on the *Candida* Genome Database (Ashburner *et al.*, 2000, Skrzypek *et al.*, 2017) and GSEA (Subramanian *et al.*, 2005). Venn diagrams were generated from lists using the online tool from the Gent bioinformatics group (http://bioinformatics.psb.ugent.be/webtools/Venn/).

7.3 Results

7.3.1 ChIP protocol optimisation

The protocol for the initial ChIP required optimisation for use in *C. albicans*, as some of the protocols were adapted from *S. cerevisiae*. Growing, fixing and spheroplasting were performed as described by Haran *et al.* in 2014 (Haran *et al.*, 2014), however the conditions for MNase digestion required optimisation. The final conditions used for MNase digestion were 7 units of MNase and 4 minutes of digestion at 37 °C. Figure 7.1 (A) shows MNase digestions of AHY940 cells with 5 units and 7 units of MNase for 4 minutes (after MNase digestion stage but before immunoprecipitation, to confirm digestion). The laddering effect of bands at 150 bp increments is indicative of optimal digestion. Samples were separated in to digestion supernatant and digestion pellet samples, however it was found that as there was some amount of DNA remaining in the pellet it was best to immunoprecipitate DNA from the entire digestion (supernatant and pellet combine) in order to maximise yield. Figure 7.1 (B) shows a Tapestation trace performed by Genewiz showing digestion of *ΔΔtlo* mutant DNA after MNase digestion and immunoprecipitated with an anti-RNA Polymerase II (anti-RNAP) antibody. These protocols were then used with to the *TLO* reintegrated strains without any need for further optimisation.

7.3.2 Sequencing quality

An initial test sequencing run was performed on an AHY940 sample and a *ΔΔtlo* mutant sample which had been MNase digested and IP performed with an anti-RNAP antibody, Bioanalyzer trace in Figure 7.1 (B). This was to ensure that enough raw material of good integrity was being produced and that the sequencing company would be able to produce good quality data from the samples. These samples were successfully sequenced by Genewiz and good quality data was returned, see Appendix 5. While this test was performed by Genewiz, subsequent sequencing was performed by Novogene. Quality and yields of all other samples can be found in Appendix 5. A much larger depth of sequencing was obtained in the sequencing performed by Novogene. Sequencing was performed on two replicates of IP from each of AHY940 (anti-RNAP), *ΔΔtlo* (anti-RNAP),

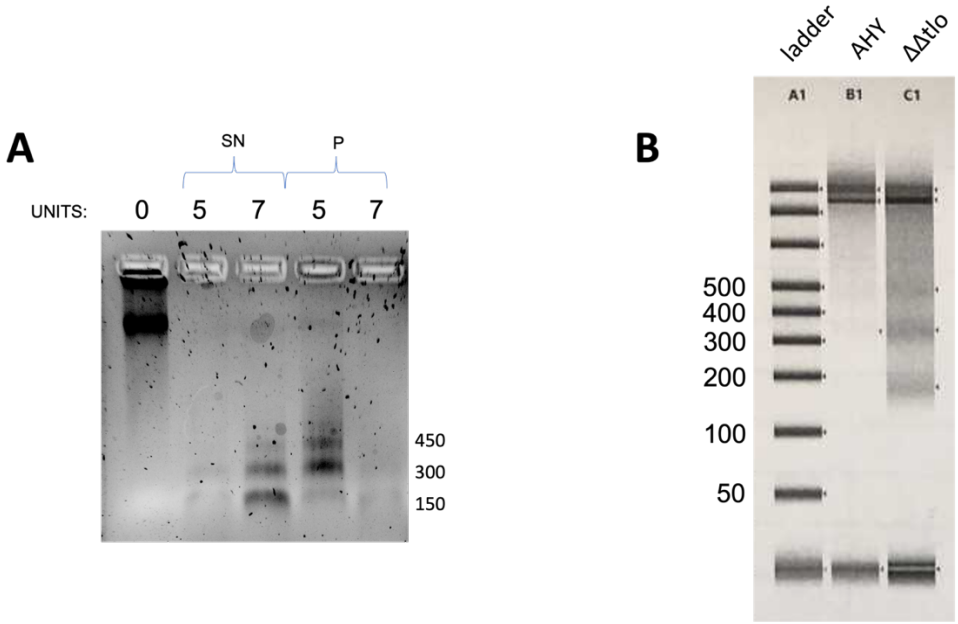


Figure 7.1 MNase digestion of DNA

The micrococcal nuclease (MNase) digestion of *C. albicans* DNA was optimised for use in the ChIP-seq protocol. (A) DNA from AHY940 strain spheroplasts subjected to MNase digestion with either 0, 5 or 7 units of MNase for four minutes at 37 °C. SN indicates DNA obtained from the supernatant of the digestion after pelleting cell debris, P indicates DNA that was extracted from the remining pellet. Lane SN 7 displays characteristic digestion of DNA, with enrichment of the 150 bp band and laddering at approx. 300 and 450 bp. (B) Tapestation read-out provided by Genewiz showing the sizes of DNA fragments sent to them for sequencing. Both strains show correct digestion, with enrichment of bands at 150, 300 and 450 bp in size.

ΔΔtlo::P_{TET}TLOα1 (anti-HA), *ΔΔtlo::P_{TET}TLOβ2* (anti-HA) and *ΔΔtlo::P_{TET}TLOγ11* (anti-HA), and a single control sample of each strain.

6.3.2.1 Alignment and peak calling

Quality statistics of the raw files of the paired end sequencing runs returned from the sequencing companies can be found in Appendix X. These files were filtered for quality and to remove sequencing adapters, then the files were aligned to the reference genome, resulting in a set of BAM files. For each sample two IP samples were sequenced along with one input control sample (no selection by immunoprecipitation). Peak calling used the control sample to identify areas of enrichment in the IP samples. Broad peaks were called for the anti-RNAP samples (AHY940 and *ΔΔtlo* mutant), and narrow peaks for the anti-HA *TLO* samples. Peaks were called individually in each IP versus the control to compare the IP samples for reproducibility, then IP reads were combined for peak calling for subsequent comparison to other strains.

The difference in sequencing depth between Genewiz and Novogene created some disparity between the number of peaks called for each IP in the anti-RNA Polymerase II samples. The first IP for each of AHY940 and the *ΔΔtlo* mutant were sequenced to a lower depth by Genewiz and these were peak called using a control alignment generated from Novogene with much higher depth. While the MACS2 peak caller does scale for size of the data sets there was still a difference in the number of peaks called. Table 7.1 details the numbers of peaks called for each replicate and the percentage of overlapping peaks between them. The percentage of overlapping peaks is a de facto measurement of reproducibility, ideally either 75% of all peaks found in one sample should overlap with the other, or 80% of the top 40% of peaks in one replicate should overlap with the peaks found in the other. All samples met these criteria, with the exception of the anti-HA *TLOγ11* strain (72% of all peaks overlapped, 75% of top 40 in replicate 1 overlapped with all peaks in replicate 2).

7.3.3 Investigation of RNA Polymerase II localisation in AHY940 and ΔΔtlo *mutant Candida albicans strains*

7.3.3.1 Peak calling, visualisation, and localisation

In order to visualise the binding of RNAP across the genome, the broad peak files generated by the peak caller, MACS2, were converted to BED6 format, to allow import into the Integrated Genome Browser (IGB). This gives a visual representation of the

	AHY940 (broad)	*ΔΔtlo* (broad)	*TLOα1* (narrow)	*TLOβ2* (narrow)	*TLOγ11* (narrow)
Replicate 1	338	39	728	3453	1382
Replicate 2	1009	463	387	2297	2536
% intersect	88%	80%	83%*	88%	75%*
Combined	850	447	635	2593	2974

Table 7.1 ChIP-seq peak calling reproducibility

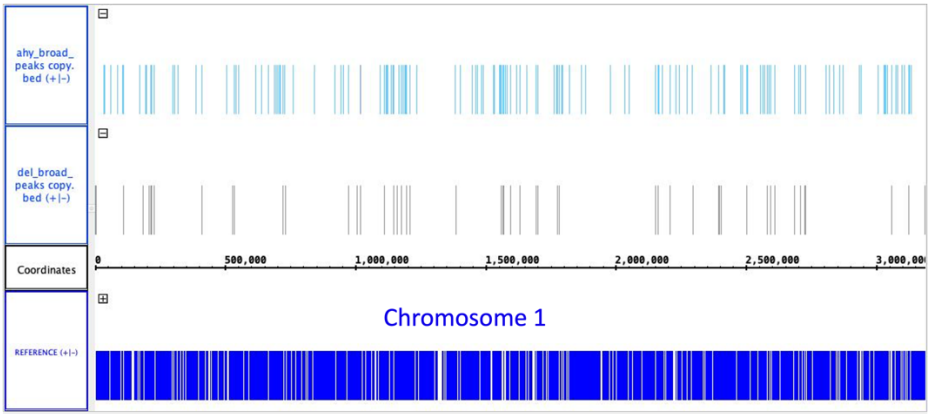
The number of peaks called for each replicate sample ChIP are detailed above. RNAP ChIP samples were called for broad peaks, which cover larger areas, while anti-HA (anti-Tlo) samples were called for narrow peaks, which are more specific regions, typically used for analysing transcription factor binding. The percentage intersect was determined using the Bedtools suite (bedtools intersect) to determine how many peaks overlapped between samples. The ideal is that over 75% of peaks are shared between replicates, or that over 80% of the top 40% of peaks in one replicate are shared (if this calculation was performed then the figure is marked with an asterisk here). After analysis of reproducibility, the replicates were combined. This was done by running the peak caller again this time using the two input samples at once. The number of peaks called by this method is also detailed above also.

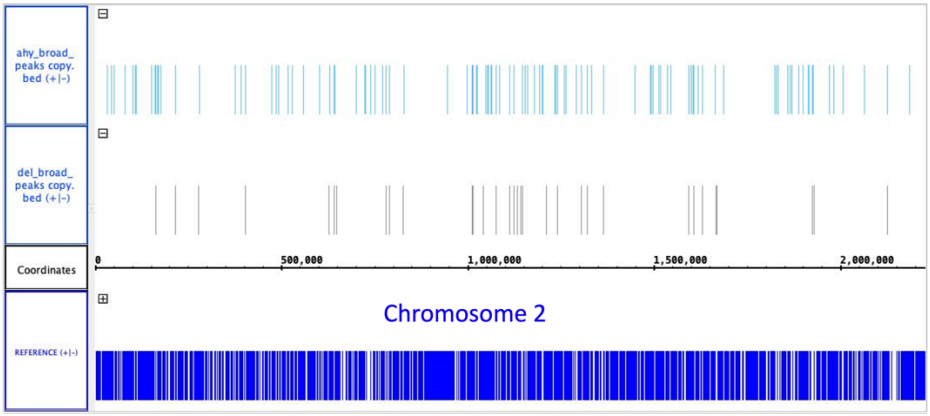
difference in RNAP interaction with the genome in the *ΔΔtlo* mutant strain compared to in AHY940, Figure 7.2.

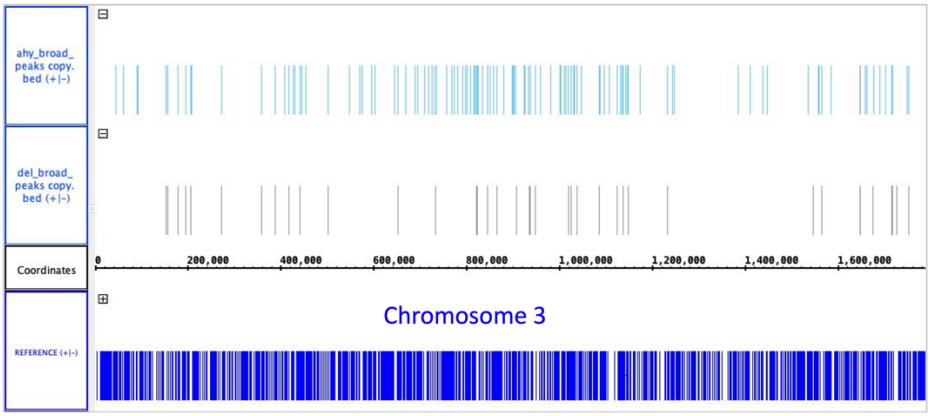
The broad peaks identified in the anti-RNAP AHY940 and *ΔΔtlo* ChIP samples were cross-referenced with the reference genome in order to determine where the RNAP was localising in the mutant and WT strains. First, the intersection of the peaks with repeat sequences, such as long terminal repeats, centromeres or major repeat sequence regions, were examined, see Figure 7.3 (A). The RNAP was found over more of these repeat regions in the AHY940 strain(n = 15) compared to the *ΔΔtlo* strain (n = 9). The majority of repeat sequence interactions identified in the *ΔΔtlo* strain were shared with the AHY940 strain. The position of RNAP in relation to genes was also analysed, over the ORF, transcription start site (TSS) or promoter (500 bp upstream of TSS). In general, RNAP was found to localise to far fewer genes in the null mutant in comparison with the WT. In the AHY940 WT strain RNAP interacted in some capacity with 1,138 unique genes; it was localised at the ORF in 989 of these cases, at the TSS in 543 and at the promoter region in 710 of these genes. In the *ΔΔtlo* mutant RNAP was found to interact with 502 unique genes, it was localised to the ORF of 422 of the3se genes, at the TSS of 222 and at the promoter region of 296 of these genes, see Figure 7.3 (A-F). The numbers of genes that RNAP interacted with in AHY940 was compared to that of the *ΔΔtlo* mutant and can be seen in Figure 7.3. In all cases it was seen that RNAP interacted with the genome more in AHY940 than in the *ΔΔtlo* mutant.

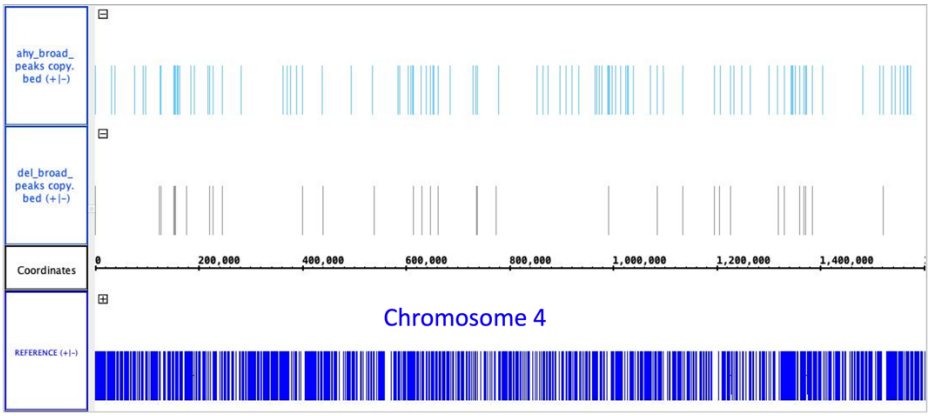
7.3.3.2 Gene Ontology analysis

The lists of genes found to be bound in any capacity by RNAP in the AHY940 and the *ΔΔtlo* mutant were subjected to GO analysis to determine if any GO terms were enriched in these datasets. When the list of genes bound by RNAP was initially subjected to GO analysis, a large number of tRNA genes were identified. When these tRNA genes were excluded from the analysis, the GO term enrichments identified were more diverse. Analysis of all genes bound by RNAP in AHY940 found enrichment for biological process GO terms; transmembrane transport (GO:0008643), carbohydrate transport (GO:0055085) and localisation (GO:0051179), Figure 7.4 (A), and of all genes bound by RNAP in the *ΔΔtlo* mutant the most highly represented biological process GO terms were mitotic cell cycle (GO:0000278) and cell cycle (GO:0007049), Figure 7.4 (B). When the dataset of genes bound by RNAP in both strains, i.e. AHY ∩ TLO_null in Figure 7.3 (E), (excluding tRNA genes) was subjected to GO analysis, the only term found to be commonly enriched was cell cycle (GO:0007049). The list of genes only bound by RNAP

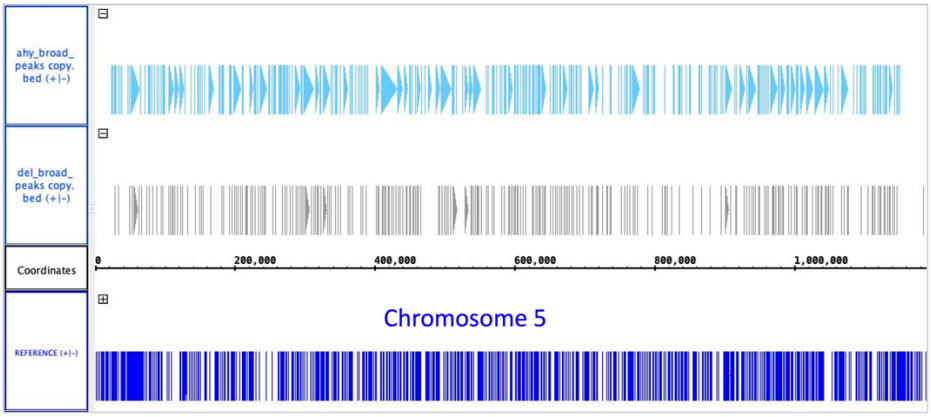


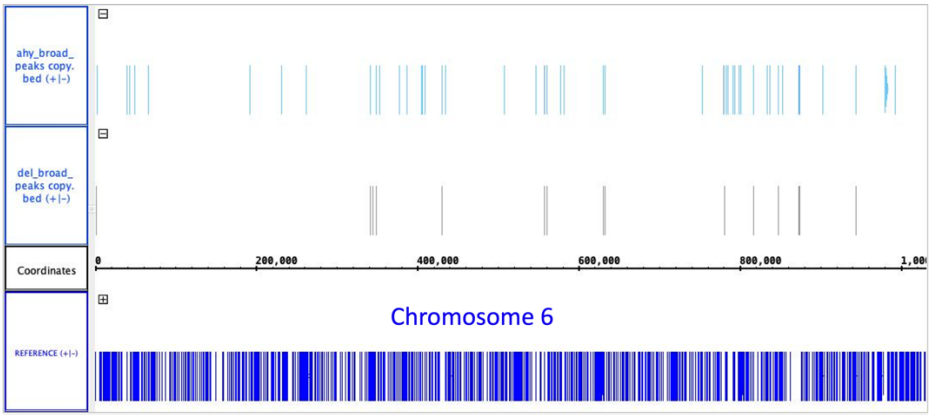


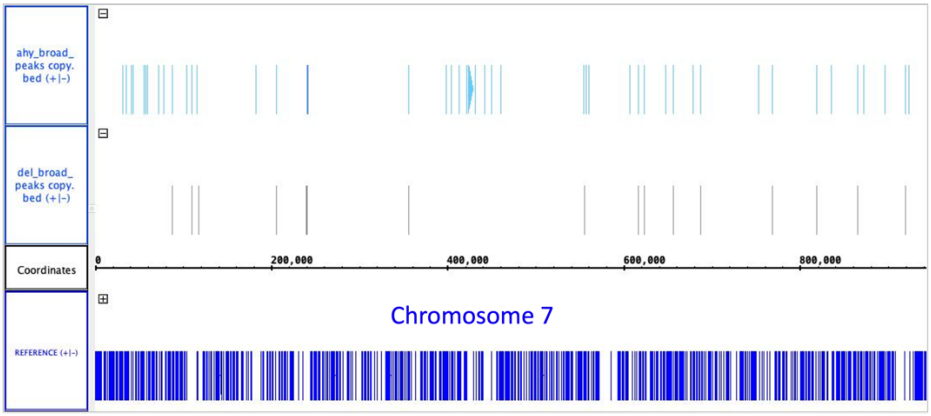


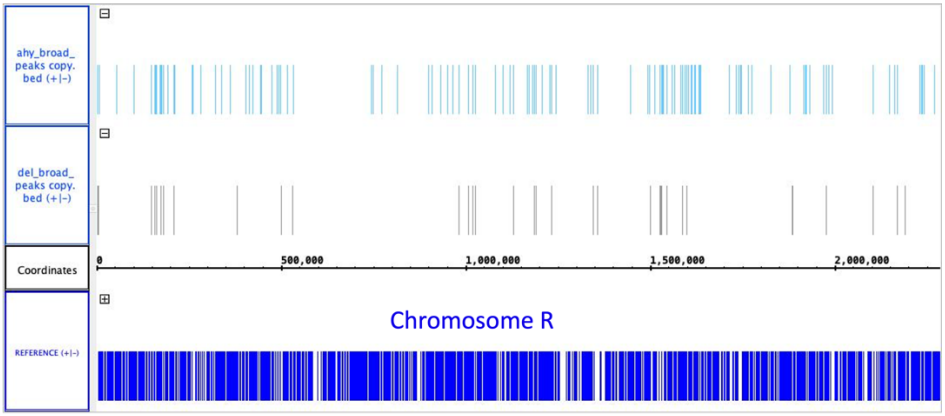


7.2 Visualisation of RNAP binding across chromosomes (cont.)





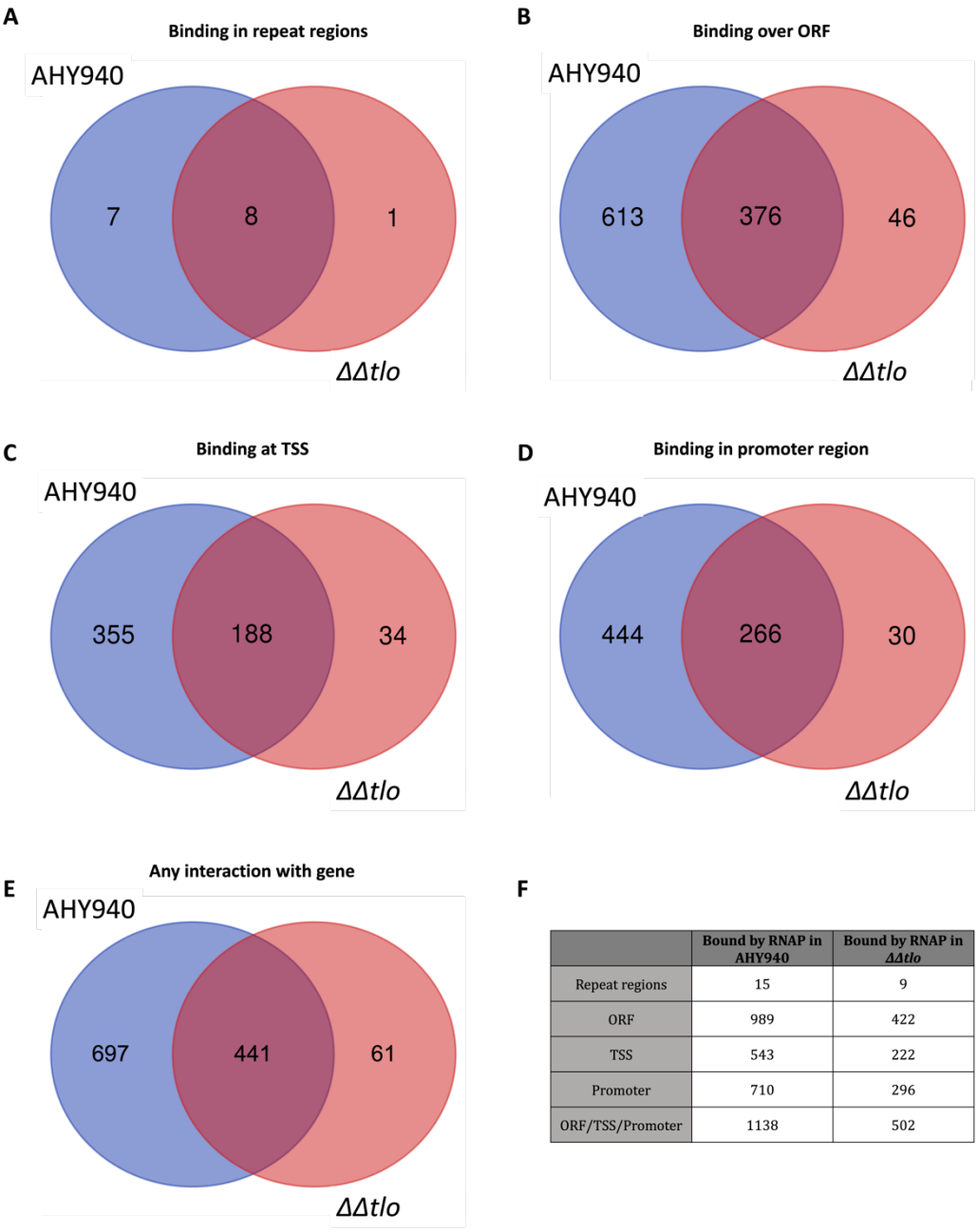




7.2 Visualisation of RNAP binding across chromosomes (cont.)

7.2 Visualisation of RNAP binding across chromosomes (cont.)

The broad peaks called by MACS2 were converted into BED6 format and inputted into the Integrated Genome Browser (IGB) to visualised the locations of peaks over the chromosomes. Peaks are coloured by score, with darker colours indicating a stronger peak score. The peaks from the anti-RNAP ChIP-seq experiment in AHY940 are on the top lane (light blue), and peaks from the anti-RNAP ChIP-seq experiment in the ΔΔtlo mutant are in the middle lane (grey). The bottom lane (dark blue) corresponds to the reference genome and the genetic features along each chromosome, with the chromosome names indicated for each panel.



F

	Bound by RNAP in AHY940	Bound by RNAP in *ΔΔtlo*
Repeat regions	15	9
ORF	989	422
TSS	543	222
Promoter	710	296
ORF/TSS/Promoter	1138	502

Figure 7.3 Location of RNA Polymerase II interaction sites in AHY940 and *ΔΔtlo* mutant

Lists of peaks from the RNAP ChIP in both AHY940 and the *ΔΔtlo* mutant were analysed with the bedtools suite to determine if the peaks overlapped with any features in the *C. albicans* reference genome. (A) Repetitive regions of the genome that were intersected by an RNAP peak. (B) ORFs that were intersected by an RNAP peak. (C) Transcription start sites (TSS) (a single base pair at the start of each gene) that were intersected by peaks. (D) Promoter regions (500 bp upstream of TSS) that were intersected by RNAP peaks. (E) All unique genes that were intersected either over promoter region, TSS or ORF by RNAP peaks. (F) Numbers of genes intersected in each category by RNAP in either AHY940 or the *ΔΔtlo* mutant.

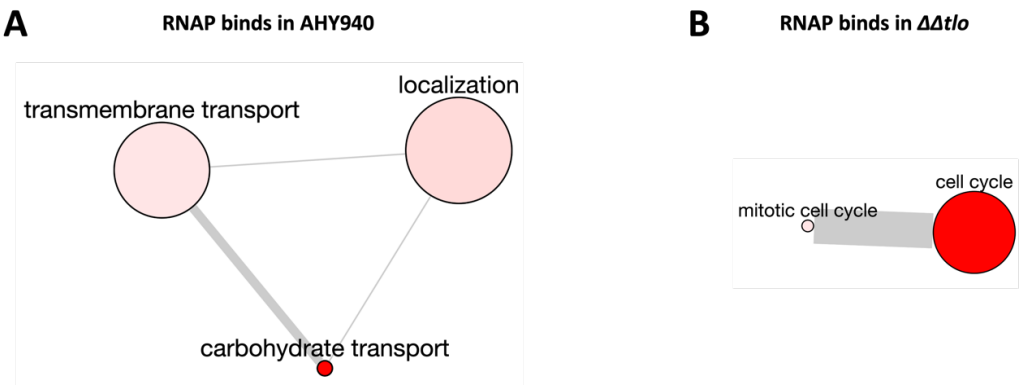


Figure 7.4 GO analysis of genes intersected by RNAP peaks in AHY940 the *ΔΔtlo* mutant

The genes intersected (at the promoter, TSS or in the ORF) by RNAP peaks in AHY940 (A) and in the *ΔΔtlo* mutant (B) were subjected to GO analysis. Biological process GO terms that were found to be enriched in these lists are displayed, with lighter coloured bubbles indicating more significance. The size of the bubbles indicates frequency of the GO term in the underlying database. Similar GO terms are joined by edges (lines) and the width of the line indicates the degree of similarity (thicker = more similar). Data was visualised using Revigo.

in AHY940 was found to be enriched for biological process GO terms carbohydrate transport (GO:0055085) and transmembrane transport (GO:0008643) only. There were no biological process GO terms found to be enriched in the list of genes bound by RNAP only in the *ΔΔtlo* mutant.

7.3.3.3 Gene Set Enrichment Analysis

The datasets of genes bound by RNAP in AHY940 and in the *ΔΔtlo* mutant, along with the score of the peak intersecting the gene, were subjected to GSEA. The list of genes and scores in the *ΔΔtlo* mutant were subtracted from the list of genes and scores in AHY940, creating a list where genes with a positive value were more enriched in AHY940 and those with a minus value were more enriched in the *ΔΔtlo* mutant. This analysis found that 638 gene sets of 703 were enriched in AHY940, 17 of them statistically significantly, while only 65 were enriched in the *ΔΔtlo* mutant, with only 4 of these being statistically significant. There was found to be enrichment of all of the gene sets corresponding to chromosome 5 in the AHY940 strain, with Figure 7.5 (A&B) showing enrichment plots for the "CA21CHR5X3" and "CA21CHR5X4" gene sets which represent the 350,000-650,000 bp and 550,000-850,000 bp positions on this chromosome respectively. All other chromosome 5 sets were significantly enriched in AHY940, but because they fell outside of the top 20 enrichments, plots were not generated by the GSEA software. Gene sets "MED31_DN", which contains genes downregulated in a ΔΔmed31 mutant, Figure 7.5 (C) and "UPC2_UP", Figure 7.5 (D), which contains genes upregulated in a ΔΔupc2 mutant, were also found to be significantly enriched in AHY940. Gene sets related to carbohydrate metabolism were also significantly enriched in the AHY940 RNAP bound gene set, including "CARBOHYDRATE METABOLIC PROCESS_BIO" (biological process GO:0005975) and "CELLULAR CARBOHYDRATE METABOLIC PROCESS_BIO" (biological process GO:0044262). Gene sets significantly enriched in the ΔΔtlo mutant included "CA21CHRRX1", containing genes found in the first 250,000 bp of chromosome R, "CA21CHR1X16", containing genes found between positions 2,950,000-3,250,000 bp of chromosome 1, as well as the gene set "SECRETORY PATHWAY_BIO", Figure 7.5 (E), which contains the genes under the secretory pathway biological GO term (GO:0045045), and the "ERV14_BIND" set, Figure 7.5 (F), which contains genes identified in a ChIP-chip microarray as being binding targets of the Erv14 protein in *S. cerevisiae*.

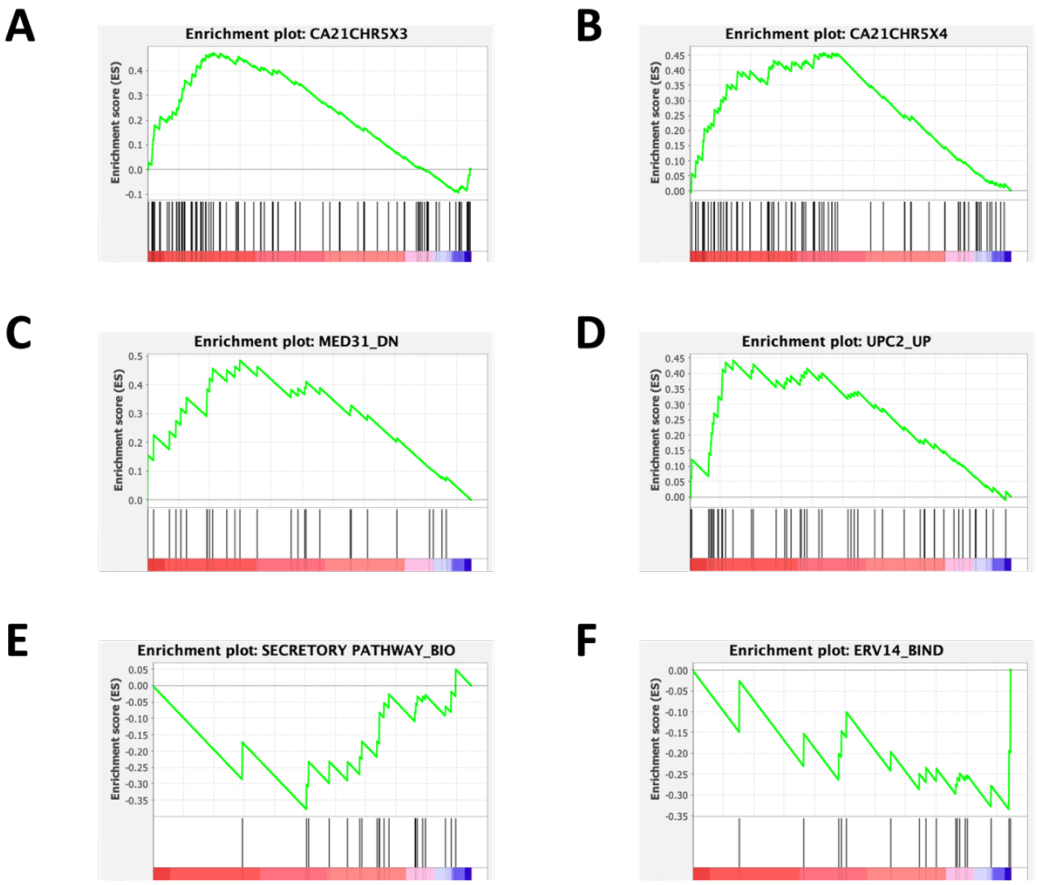


Figure 7.5 Enrichment plots comparing gene sets enriched in genes intersected by RNAP peaks in AHY940 vs. the *ΔΔtlo* mutant

Enrichment plots were generated using GSEA software, comparing the lists of genes intersected by RNAP peaks in AHY940 and the *ΔΔtlo* mutant (ranked by peak score). Plots depict the enrichment score (top, green) for the genes across the ranked list, with the black bars indicating where each gene in the set falls in the list. Skew to the left here indicates enrichment in AHY940, as does the enrichment score being a positive number, and skew to the right indicates enrichment in the *ΔΔtlo* mutant, as does a negative enrichment score. (A) Enrichment plot for gene set located on chromosome 5 (350,000-650,000 bp). (B) Enrichment plot for gene set located on chromosome 5 (550,000-850,000 bp). (C) Enrichment plot for gene set downregulated in a *ΔΔmed31* mutant. (D) Enrichment plot for gene set upregulated in a *ΔΔupc2* mutant. (E) Enrichment plot for gene set involved in the secretory pathway biological process GO term (GO:0045045). (F) Enrichment plot for gene set bound by Erv14 (*S. cerevisiae)*.

7.3.3.4 Relationship between RNAP interaction and gene expression

To investigate the effect of RNAP binding on gene expression, the ChIP-seq data from this chapter was compared to the RNA-seq data from Chapter 6. Here, the expression levels of genes identified as being intersected by RNAP peaks were examined. The mean raw expression levels of the RNAP intersected genes in each strain were compared to the mean expression levels of all the genes not intersected by RNAP peaks in each strain. It was found that there was no significant difference in the expression levels of genes intersected compared to those not intersected by RNAP peaks in either strain (data not shown). The mean expression of genes commonly intersected by RNAP peaks in both strains was also compared, Figure 7.6 (A), and it was determined that there was no significant difference in expression of these genes. In the case of genes that were intersected by RNAP peaks in only AHY940, the mean expression level of these genes in AHY940 was compared to the mean expression level of these genes in the *ΔΔtlo* mutant, Figure 7.6 (B). Here it was seen that these genes were expressed at a significantly higher level in AHY940 compared to in the *ΔΔtlo* mutant. The same analysis was performed for genes only intersected by RNAP peaks in the *ΔΔtlo* mutant, Figure 7.6 (C), however, there was no difference in expression level detected in these genes.

Similar analysis was performed in order to determine if binding position, promoter, TSS or in the ORF had any effect on expression. In the case of genes intersected by RNAP peaks in AHY940, it was found that genes bound at either the TSS or in the ORF were expressed at a higher level than those intersected in the promoter region (500 bp upstream of TSS), Figure 7.7 (A). In the case of genes intersected by RNAP peaks in the *ΔΔtlo* mutant, no difference was found in the expression levels of genes intersected at the various regions, Figure 7.7 (B).

7.3.4 Investigation of Tlo localisation in the TLO reintroduction strains

7.3.4.1 Peak calling, visualisation and localisation

The narrow peaks identified in the anti-HA *TLO* strain ChIP samples were subjected to similar analysis as above to examine the relationship between the Tlo proteins and the genome. In order to visualise the binding of the Tlo proteins across the genome, the narrow peak files generated by the peak caller, MACS2, were converted to BED6 format, to allow import into IGB. This allowed for visualisation of Tlo interaction across the chromosomes, Figure 7.8, and visualisation of interaction at specific genes, Figure 7.9. Tlo binding over the transcription factors *TYE7,* Figure 7.9 (A), *GAL4,* Figure 7.9 (B), and

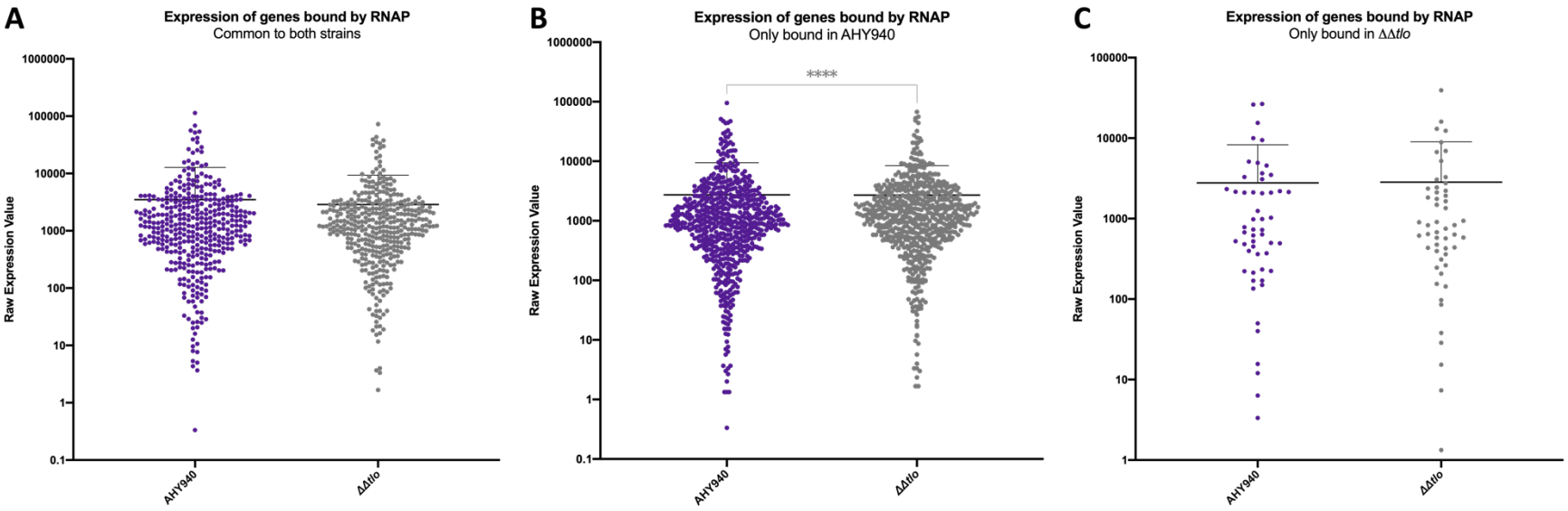


Figure 7.6 Expression of genes intersected by RNAP peaks in AHY940 and the *ΔΔtlo* mutant

The raw expression values of genes intersected by RNAP peaks in AHY940 and in the *ΔΔtlo* mutant were plotted. Wilcoxon matched-pairs signed rank tests were performed to determine significant difference between groups, bars with asterisks denote significant difference. (A) Raw expression of genes intersected by RNAP peaks in both AHY940 and the *ΔΔtlo* mutant. (B) Raw expression of genes only intersected by RNAP peaks in AHY940. (C) Raw expression of genes only intersected by RNAP peaks in the *ΔΔtlo* mutant.

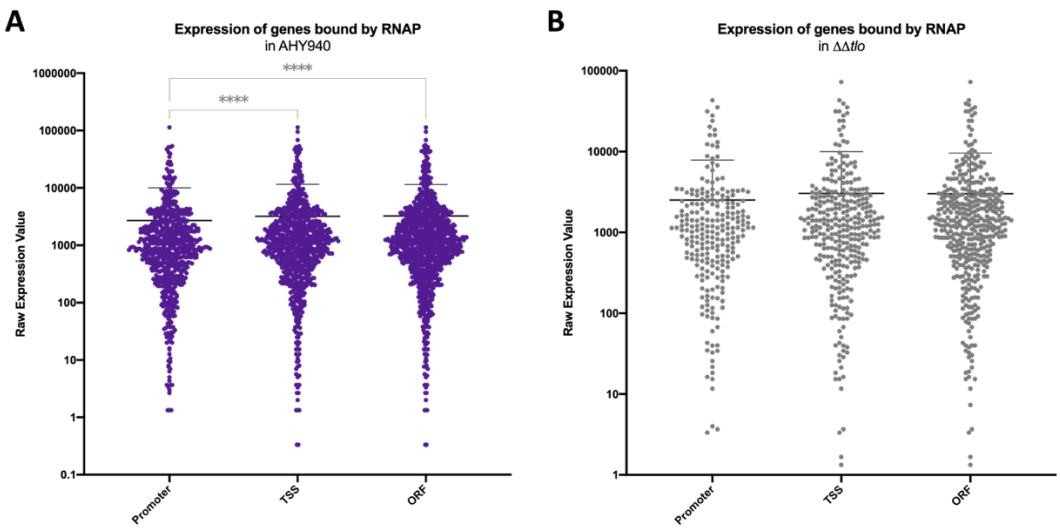


Figure 7.7 Expression of genes intersected by RNAP peaks at different sites

The raw expression values of genes intersected by RNAP peaks at the promoter, at the TSS and over the ORF were plotted for AHY940 and the *ΔΔtlo* mutant. Brown-Forsythe and Welch ANOVAs with multiple comparisons to compare the means of each group were performed to determine if differences were significant. Bars with asterisks denote significant differences between groups. (A) Expression of genes intersected by RNAP peaks in AHY940. (B) Expression of genes intersected by RNAP peaks in the *ΔΔtlo* mutant.

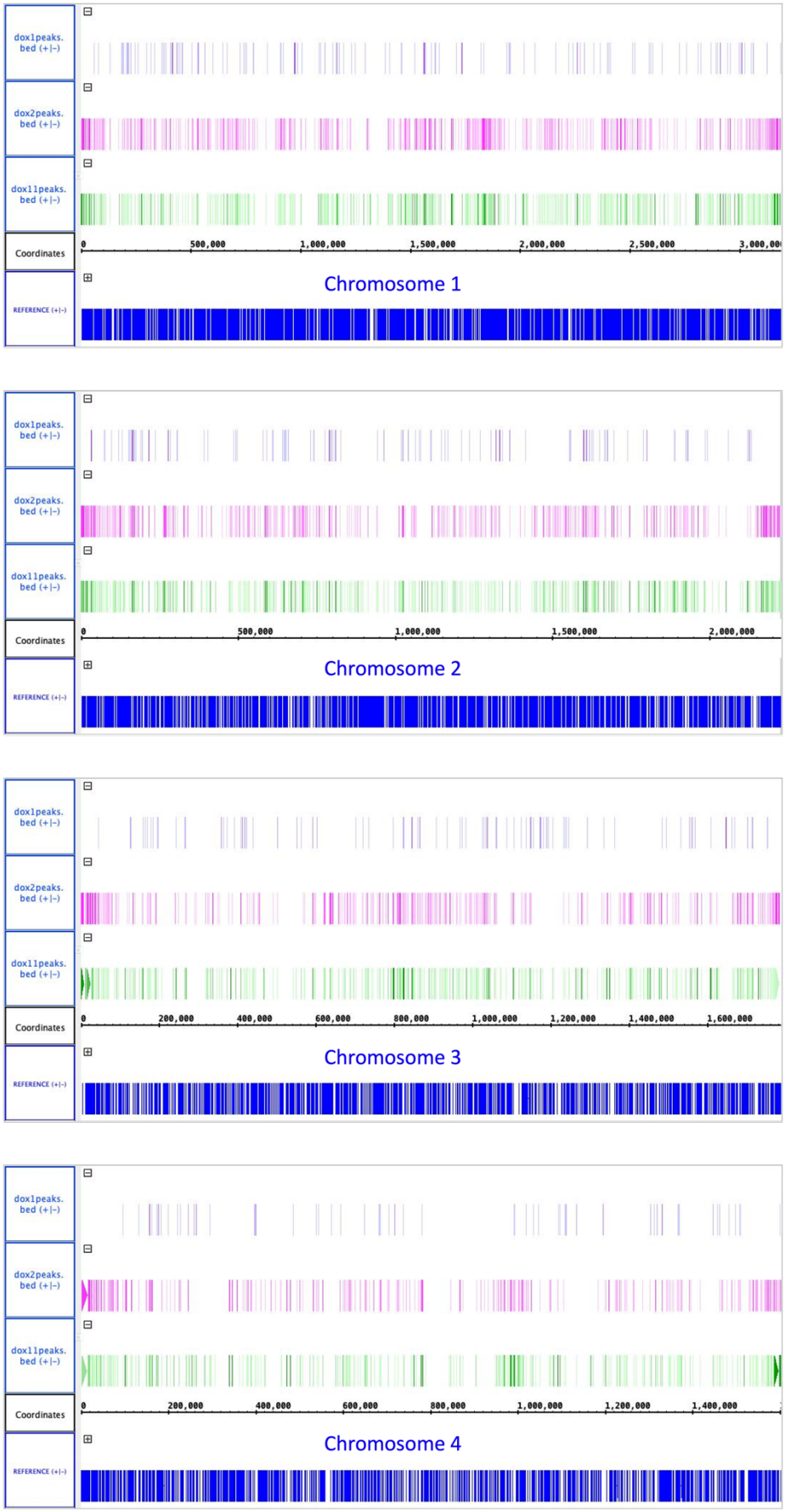


Figure 7.8 Localisation of Tlos over chromosomes (cont.)

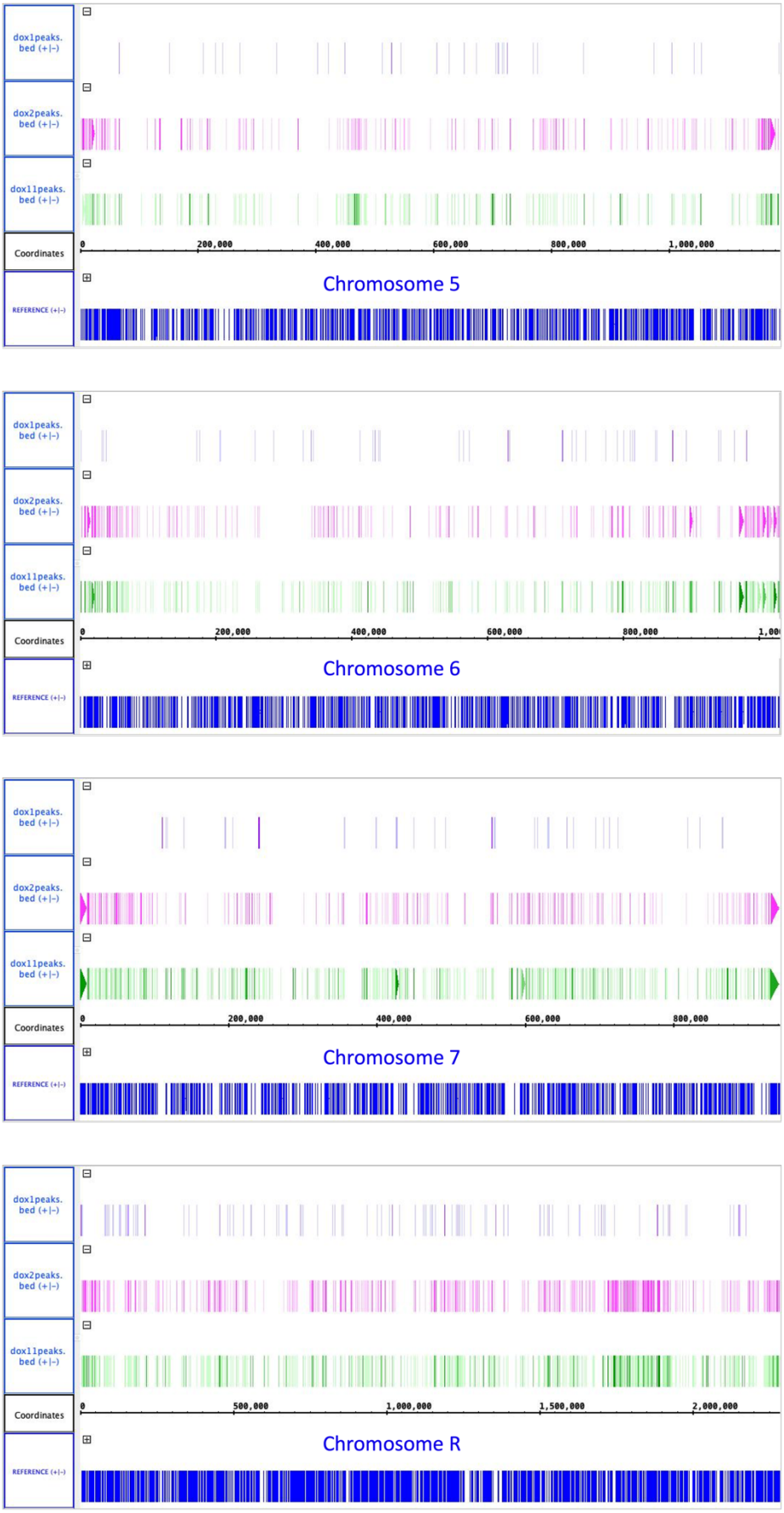


Figure 7.8 Localisation of Tlos over chromosomes (cont.)

Figure 7.8 Localisation of Tlos over chromosomes (cont.)

The narrow peaks called by MACS2 were converted into BED6 format and inputted into the Integrated Genome Browser (IGB) to visualised the locations of peaks over the chromosomes. Peaks are coloured by score, with darker colours indicating a stronger peak score. The Tloα1 peaks are found in the top lane (purple), the Tloβ2 peaks in the second lane (pink) and the Tloγ11 peaks in the third lane (green). The bottom lane (dark blue) corresponds to the reference genome and the genetic features along each chromosome, with the chromosome names indicated for each panel.

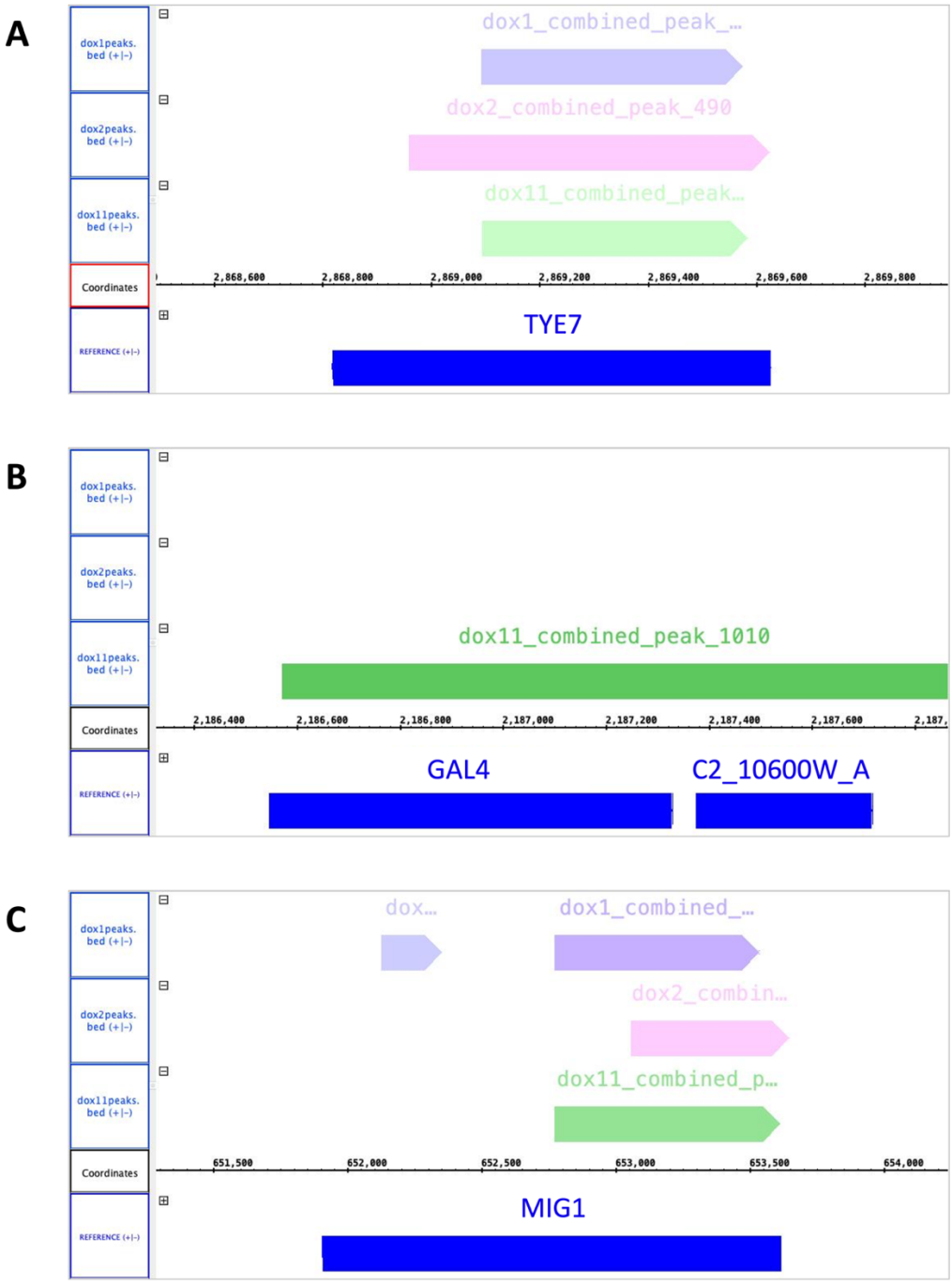


Figure 7.9 Localisation of Tlos at genes *TYE7, GAL4* and *MIG1*

The narrow peaks called by MACS2 were converted into BED6 format and inputted into the Integrated Genome Browser (IGB) to visualised the locations of peaks over the specific genes. Peaks are coloured by score, with darker colours indicating a stronger peak score. The Tloα1 peaks are found in the top lane (purple), the Tloβ2 peaks in the second lane (pink) and the Tloγ11 peaks in the third lane (green). The bottom lane (dark blue) corresponds to the reference genome and genetic features, with the gene names indicated above each feature.

MIG1, Figure 7.9 (C), were selected as they represent genes intersected by individual or multiple Tlo peaks, and were found to be relevant in later sections of this chapter.

Investigation into Tlo localisation at repetitive regions, such as centromeres, MRS, HOKs and at telomeres (LTRs), indicated that all three Tlos localised to these regions. Figure 7.10 (A) shows the binding of each Tlo to these regions as a heatmap. Tloγ11 localised at these repetitive regions most, and Tloα1 least. Representative images of peaks over a telomeric regions, the last 50 kbp of chromosome 1, and over a MRS region, MRS-R on chromosome R, can be found in Figures 7.10 (B) and (C) respectively.

The interaction of the individual Tlo proteins with genes was also investigated. Interactions of Tlo proteins with genes, either over the ORF, at the TSS or at the promoter, were identified in all three strains. Tloα1 interacted with 552 genes, of these Tloα1 interacted with 477 over the ORF, 96 at the TSS and 176 in the promoter region. Tloβ2 interacted with 2,329 genes, of these it interacted with 2,111 over the ORF, 478 at the TSS and 876 at the promoter region. Tloγ11 interacted with 2,521 genes, of these it interacted with 2,306 over the ORF, 591 at the TSS and 987 in the promoter region. Of all genes that the Tlos interact with, 275 genes were intersected by peaks of each three Tlos, Figure 7.11 (E&F). The lists of gene interactions were then compared to determine if any genes were intersected by peaks from all three Tlos. Tloα1 has the fewest interactions out of the three Tlos tested, and Tloβ2 and Tloγ11 have a similar amount of interactions. All three Tlos share interaction over the ORFs of 235 genes, Figure 7.11 (B), and of all the Tlos, Tloα1 has the fewest unique interactions here. Tloβ2 and Tloγ11 shared a large proportion of interactions over the bodies of genes. Of all types of interactions examined here, Tlos tended to interact with the bodies of genes more so than with the TSS or promoters. The Tlos shared interactions at the TSS of a small number of genes, Figure 7.11 (C), and again Tloα1 had the fewest unique interactions. Tloβ2 and Tloγ11 again shared a larger number of interactions at the TSS of genes with each other than they did with Tloα1. Interactions at the promoters of genes also follows this trend, Figure 7.11 (D), where Tloα1 shared most of its interactions with other Tlos. Tloβ2 and Tloγ11 again shared a large proportion of their interactions at promoters.

7.3.4.2 Gene Ontology analysis

Genes intersected by peaks from each Tlo were compared, and specific sets of genes were subjected to GO analysis to determine if there was any enrichment of GO terms in these groups. Lists of genes were subjected to GO analysis, and the three subcategories

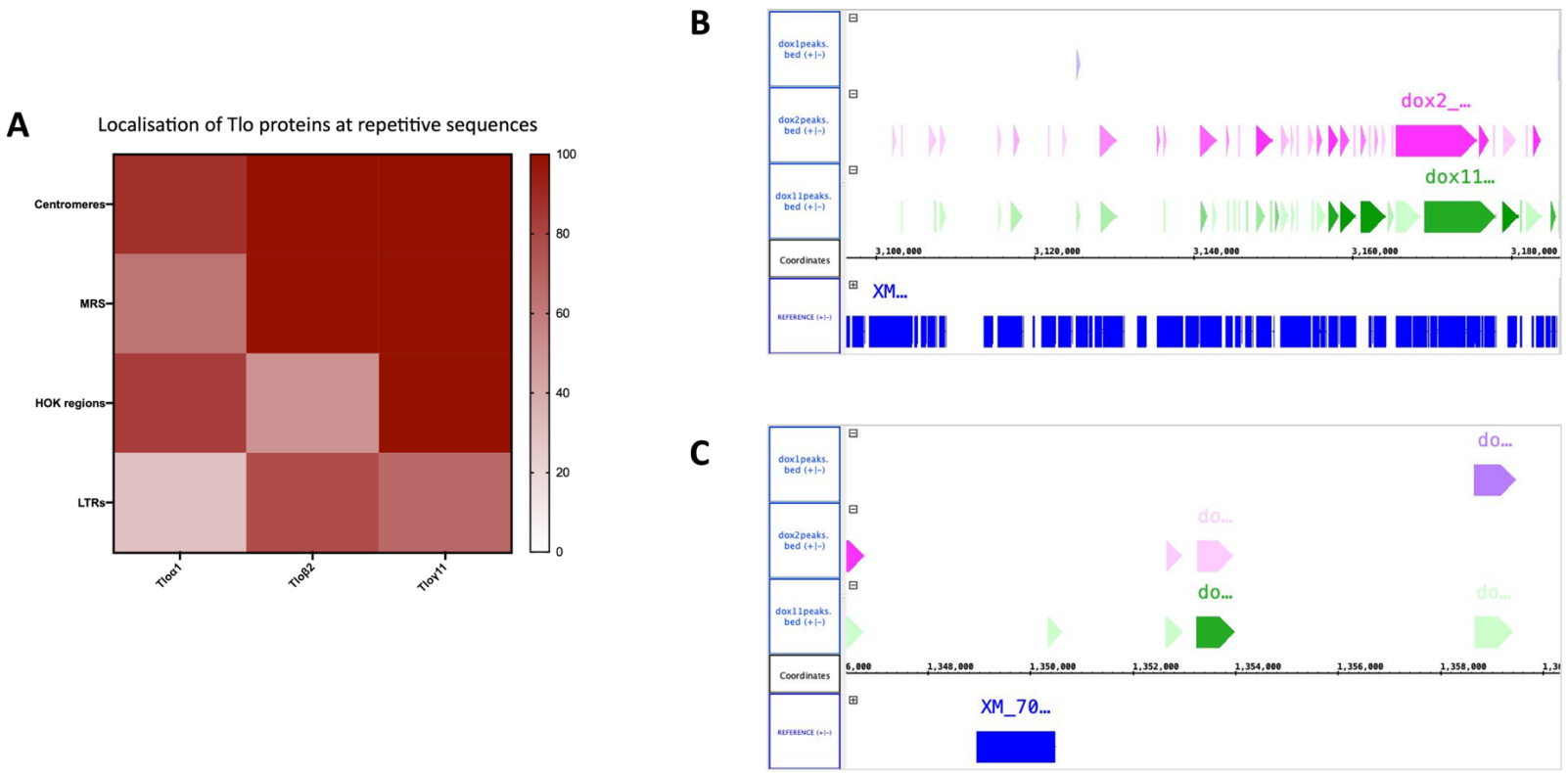
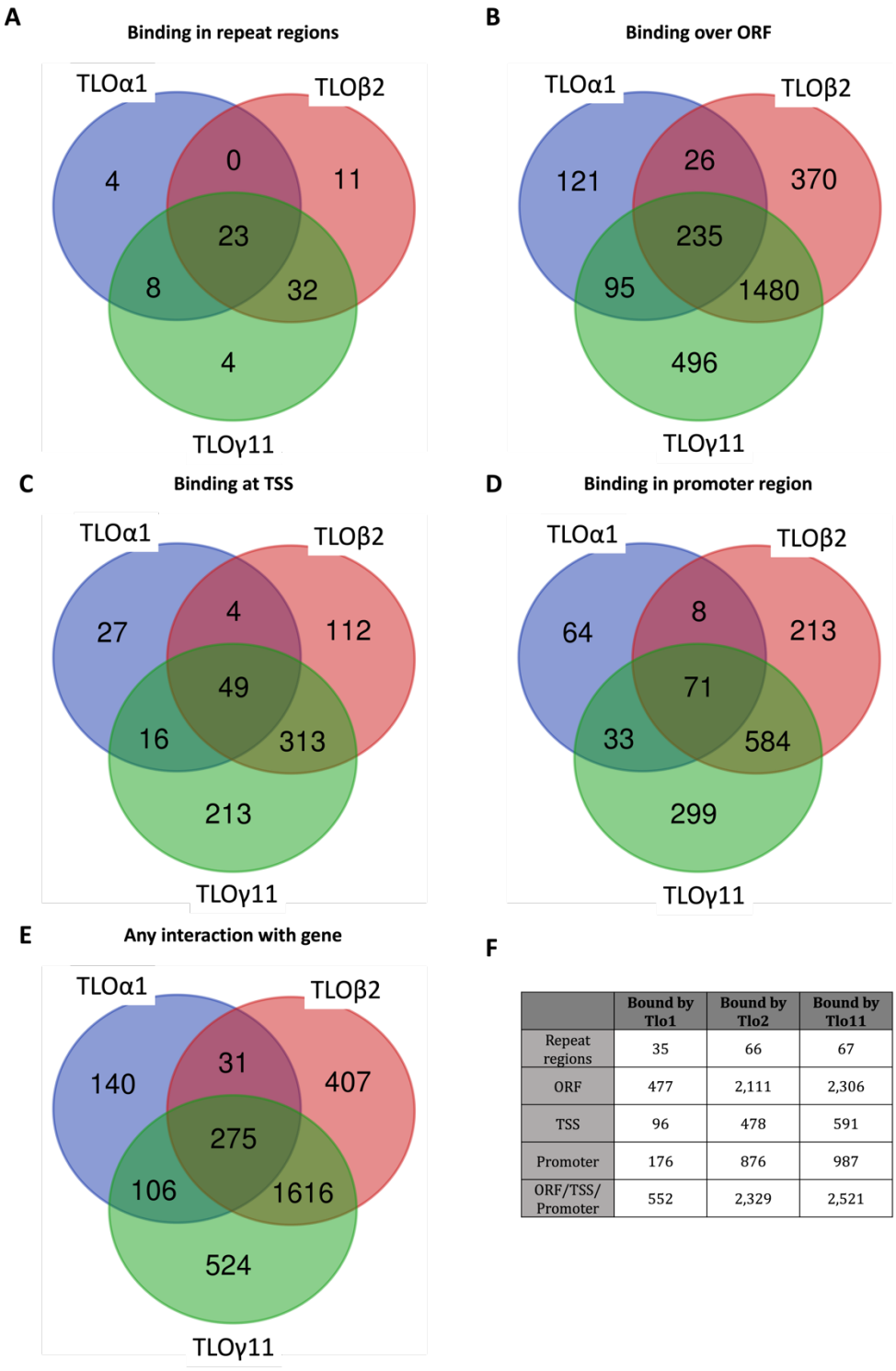


Figure 7.10 Localisation of Tlo proteins at repetitive regions

(A) Localisation of each Tlo protein to the repetitive regions of the *C. albicans* chromosome is displayed as a heatmap, where intensity is determined by the number of regions intersected by Tlo peaks out of the total number of these regions, represented as a percentage (i.e. Tloγ11 peaks intersect all centromeres = 100%), with darker colours indicating more interaction at those regions. Regions include centromeres, major repeat sequences (MRS), HOK regions and long terminal repeats (LTRs) commonly found at the telomeres. (B) Visualisation of the peaks called for each Tlo samples in IGB over the last 50 kbp of chromosome 1 (i.e. the telomere), where the Tloα1 peaks are found in the top lane (purple), the Tloβ2 peaks in the second lane (pink) and the Tloγ11 peaks in the third lane (green). Darker colour represents a higher peak score. (C) Visualisation of the peaks called for each Tlo samples in IGB over the MRS of chromosome R (MRS-R).



	Bound by Tlo1	Bound by Tlo2	Bound by Tlo11
Repeat regions	35	66	67
ORF	477	2,111	2,306
TSS	96	478	591
Promoter	176	876	987
ORF/TSS/ Promoter	552	2,329	2,521

Figure 7.11 Location of Tlo interaction sites in *TLO* reintroduction strains

Lists of peaks from the anti-HA (anti-Tlo) ChIP experiments in the *TLO* reintroduction strains were analysed with the bedtools suite to determine if the peaks overlapped with any features in the *C. albicans* reference genome. (A) Repetitive regions of the genome that were intersected by a Tlo peak. (B) ORFs that were intersected by an Tlo peak. (C) Transcription start sites (TSS) (a single base pair at the start of each gene) that were intersected by Tlo peaks. (D) Promoter regions (500 bp upstream of TSS) that were intersected by Tlo peaks. (E) All unique genes that were intersected either over promoter region, TSS or ORF by Tlo peaks. (F) Numbers of genes intersected in each category by each Tlo in the respective strains.

of GO term were examined, biological process, molecular function and cellular component. Unlike with the RNAP ChIP, the inclusion of tRNA genes did not seem to interfere with the analysis of the data, as other groups were shown to be enriched also, therefore these genes were included in the analysis. The genes that were intersected by peaks from all three Tlos were analysed, and many terms were returned as enriched in this set. In terms of biological process, Figure 7.12 (A), the most significantly enriched terms were those related to translation, e.g. cytoplasmic translation (GO:0002181), cytoplasmic translational elongation (GO:0002182) and translational elongation (GO:0006414) being the most significant, with many of the same genes being included in each of the three sets. Some other biological process terms enriched include general cellular and metabolic process terms, growth, transcription, terms related to the response to stress, filamentous growth and organelle organisation. Molecular function GO term enrichment was also investigated, Figure 7.12 (B), with mRNA binding (GO:0003729) and terms related to translation, such as triplet codon-amino acid adaptor activity (GO:0030533) being significantly enriched. Molecular function terms related to carbohydrate transmembrane transporter activity (GO:0015144) were also enriched, as well as the DNA-binding transcription factor activity term (GO:0003700). Cellular component GO terms that were enriched included terms related to the plasma membrane (GO:0005886), the mating projection tip (GO:0043332) and the cell periphery (GO:0071944), Figure 7.12 (C).

Genes that were only intersected by Tloα1 peaks (140 genes) were subjected to GO analysis. There were no enrichments found in this list for either biological process or cellular component, however, there were two terms enriched for molecular function. These terms were glycerol-3-phosphate O-acyltransferase activity (GO:0004366) and glycerone-phosphate O-acyltransferase activity (GO:0016287). This analysis was also performed for genes that were only intersected by Tloβ2 peaks (407 genes), and again there was no enrichment found for any biological process or cellular component GO terms. For molecular function, there was enrichment for the terms nucleoside-triphosphatase activity (GO:0017111) and hydrolase activity (GO:0016787).

The genes that were only intersected by Tloγ11 peaks (524 genes), were also subjected to GO analysis, and many more terms were found to be enriched in this data set than in the Tloα1 and Tloβ2 only data sets. Biological process GO terms, Figure 7.13 (A), that were significantly enriched included many related to regulation of biological processes, such as the regulation of nitrogen compound metabolic process (GO:0051171) and the

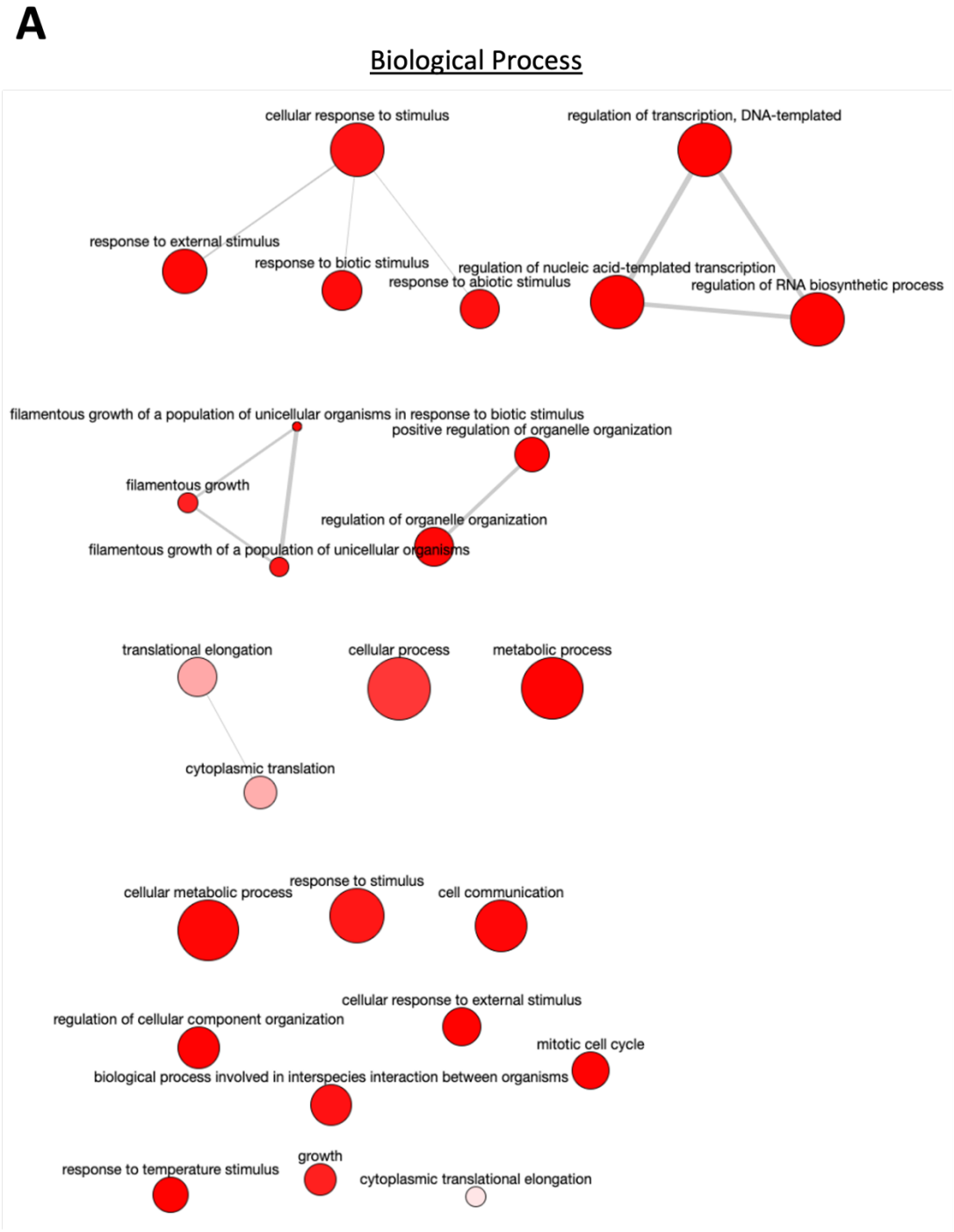


Figure 7.12 GO analysis of genes which Tloα1, Tloβ2 and Tloγ11 peaks commonly intersected (cont.)

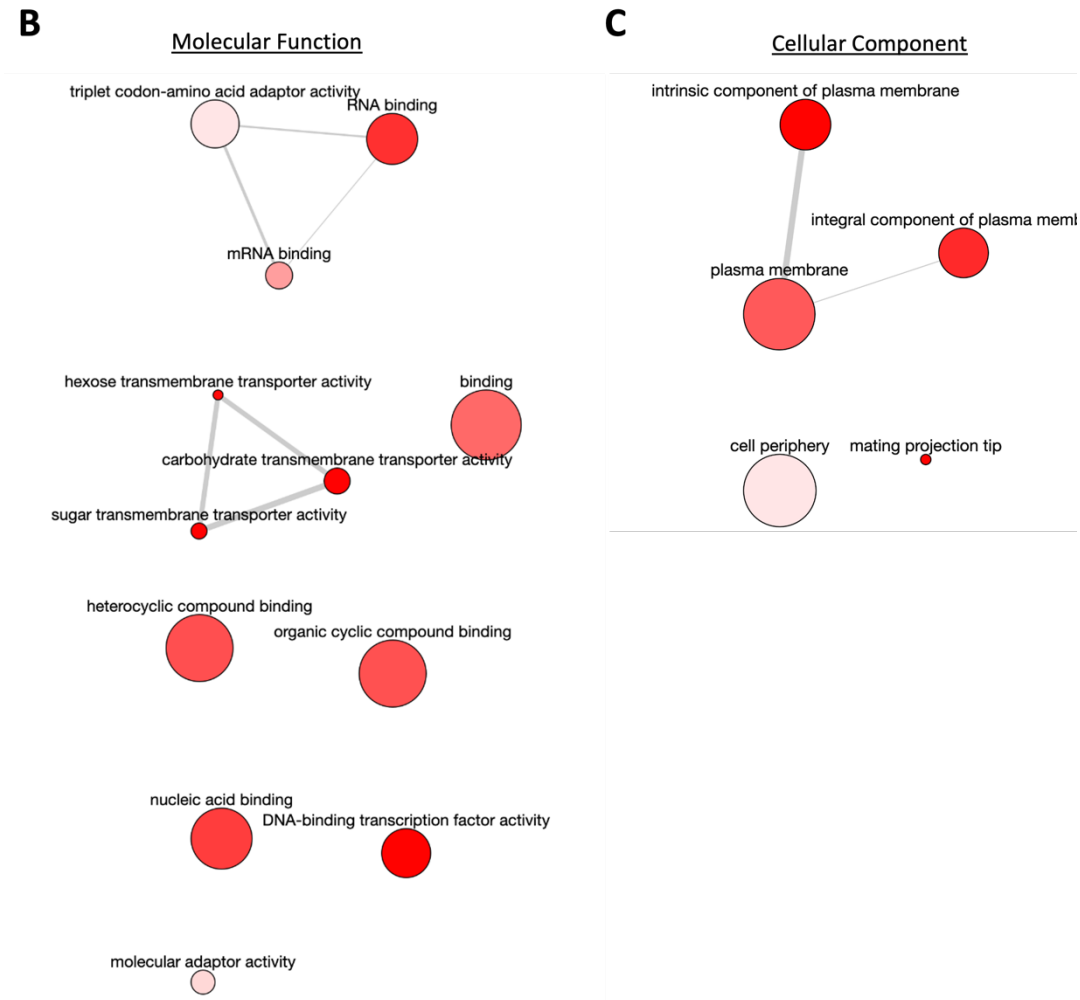


Figure 7.12 GO analysis of genes which Tloα1, Tloβ2 and Tloγ11 peaks commonly intersected (cont.)

The list of genes that all three Tlos interacted with (Tloα1 in the *ΔΔtlo::P$_{TET}$TLOα1*, Tloβ2 in the *ΔΔtlo::P$_{TET}$TLOβ2* and Tloγ11 in the *ΔΔtlo::P$_{TET}$TLOγ11*) were subjected to GO analysis. (A) Biological process GO terms. (B) Molecular function GO terms. (C) Cellular component GO terms.

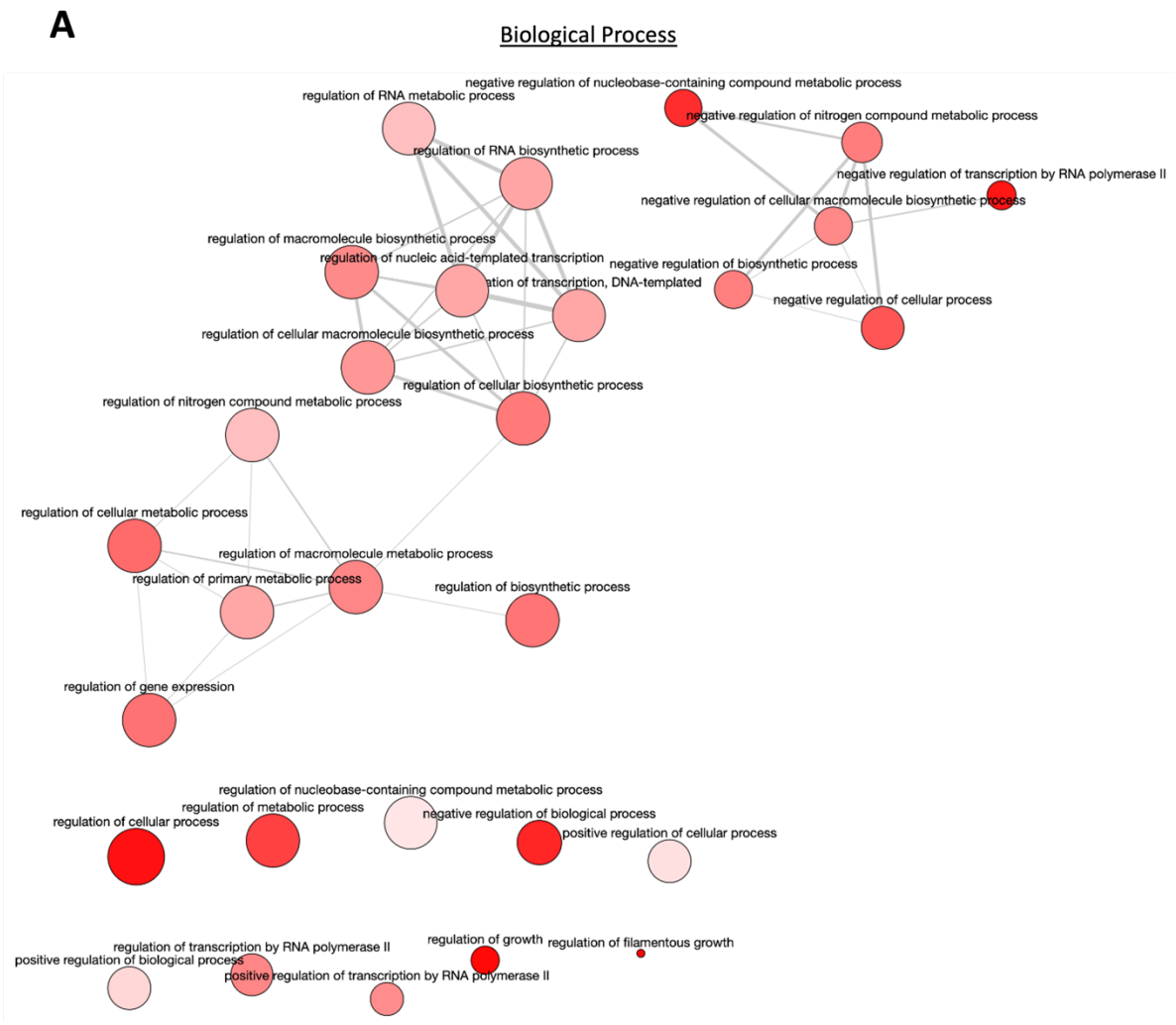


Figure 7.13 GO analysis of genes which Tloγ11 peaks intersected (cont.)

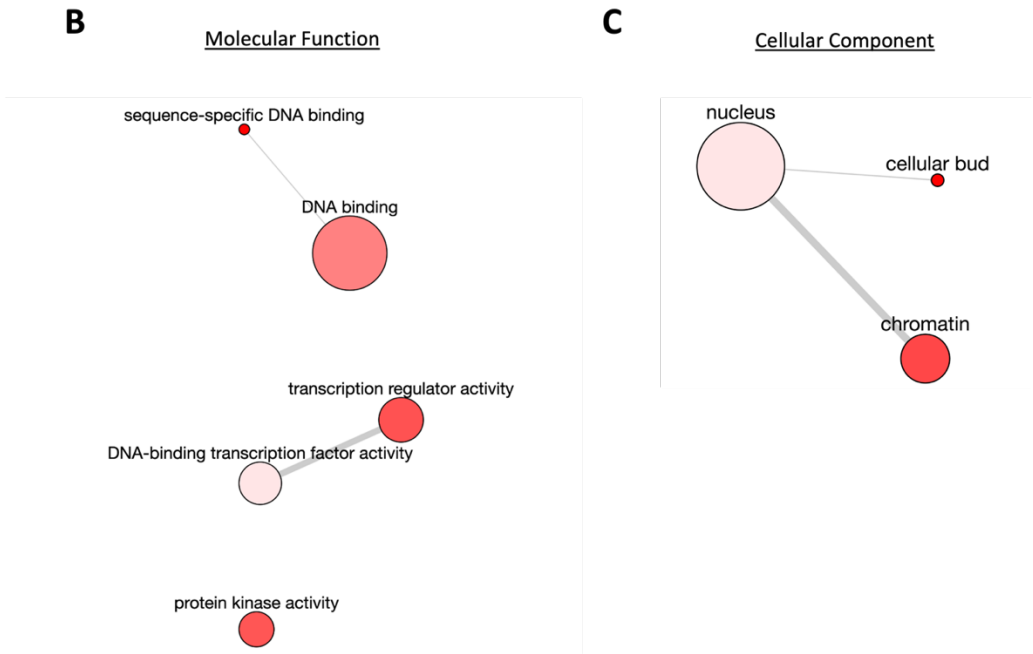


Figure 7.13 GO analysis of genes which Tloγ11 peaks intersected (cont.)

The list of genes that that Tloγ11 interacted with were subjected to GO analysis. (A) Biological process GO terms. (B) Molecular function GO terms. (C) Cellular component GO terms.

regulation of nucleic acid-templated transcription (GO:1903506). Molecular function GO terms that were enriched included DNA-binding transcription factor activity (GO:0003700) and protein kinase activity (GO:0004672), Figure 7.13 (B). Cellular component GO terms that were enriched included nucleus (GO:0005634), chromatin (GO:0000785) and cellular bud (GO:0005933), Figure 7.13 (C).

7.3.4.3 Gene Set Enrichment Analysis

The lists of genes that were intersected by peaks in each strain were subjected to gene set analysis to determine if there were any specific function or trends in the gene that each Tlo interacted with. Lists consisted of genes that were intersected by peaks, ranked by the score of the peak intersecting it (from the peak file generated by the peak caller). Lists only contained positive values. These ranked lists were then submitted to GSEA and the pre-ranked tool was used to determine enrichments.

For the Tloα1 list, there were 303 gene sets enriched, with 24 of these being significantly enriched (p =< 0.05). Top enriched sets included the gene set located at positions 1 – 250,000 bp on chromosome 2; "CA21CHR2X1", the "HS_UP" gene set which contains genes that are upregulated when the cell is exposed to heat shock, and "GAL4_UP" which contains genes upregulated in a *ΔΔgal4* mutant *C. albicans* strain, Figure 7.14 (A-C). The "MED31_DN" which contains genes downregulated in a *ΔΔmed31* mutant was also significantly enriched. There was also enrichment of the "WHITE_UP" gene set, containing genes upregulated in white cells, however this enrichment fell just outside the significance range. Other gene sets were significantly enriched such as those related to lipid and amine metabolic processes, Figure 7.14 (D), as well as the "CDC5_9H_UP" set, containing genes upregulated after 9 hours of repression of the Cdc5 Polo-like kinase.

In the case of genes that Tloβ2 interacted with, there were 1,899 gene sets enriched, with 101 of these being statistically significant. Some of the top enrichments included gene sets located at the ends of chromosomes 3, 5 and 7, Figure 7.15 (A). There was also enrichment of gene sets related to carbohydrate metabolism, including "GLUCOSE CATABOLIC PROCESS_BIO" (genes in GO:0006007), "ALCOHOL CATABOLIC PROCESS_BIO" (genes in GO:0046164) and "CELLULAR CARBOHYDRATE CATABOLIC PROCESS_BIO" (genes in GO:0044275), Figure 7.15 (B). Gene sets related to defence mechanisms were also significantly enriched, such as "RESPONSE TO HOST DEFENSES_BIO" (genes in GO:0052200) and "INDUCTION BY SYMBIONT OF HOST

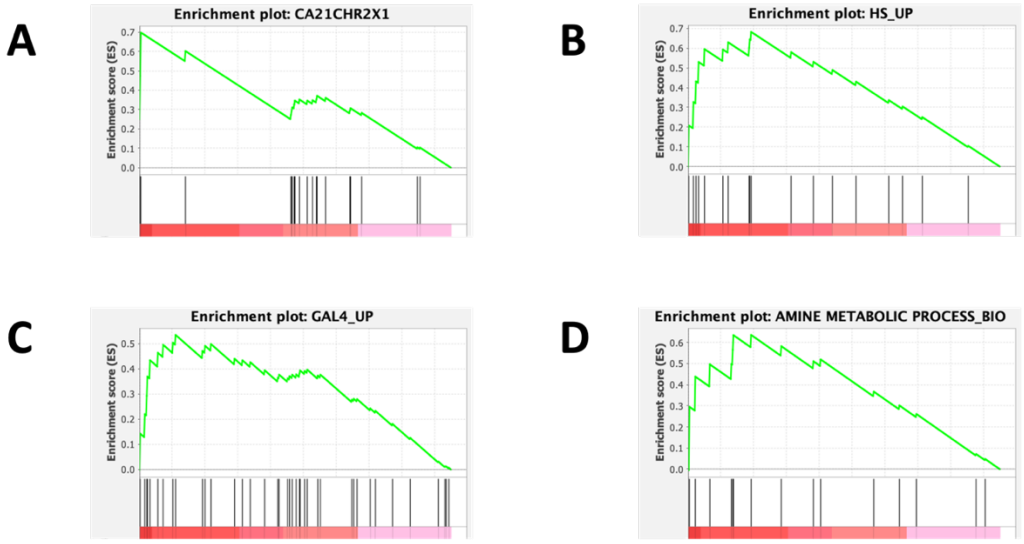


Figure 7.14 Enrichment plots of gene sets enriched in genes which Tloα1 peaks intersected

Enrichment plots were generated using GSEA software using the list of genes intersected by a Tloα1 peaks as the input (ranked by peak score). Plots depict the enrichment score (top, green) for the genes across the ranked list, with the black bars indicating where each gene in the set falls in the list. (A) Enrichment plot for the "CA21CHRX1" gene set, between positions 1-250,000 bp of chromosome 2. (B) Enrichment plot for the "HS_UP" gene set. (C) Enrichment plot for the "GAL4_UP" gene set. (D) Enrichment plot for the "AMINE METABOLIC PROCESS_BIO" gene set (GO:0009308).

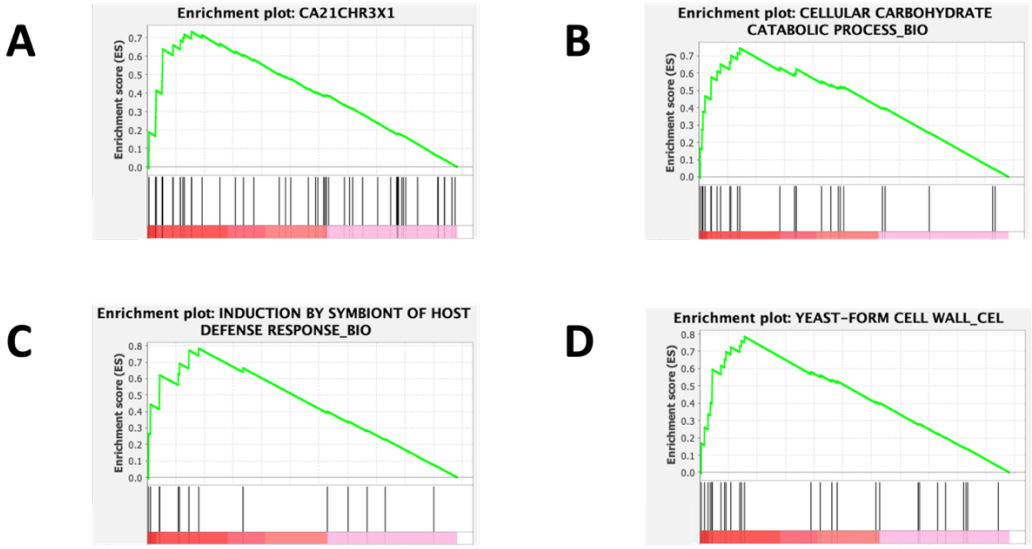


Figure 7.15 Enrichment plots of gene sets enriched in genes which Tloβ2 peaks intersected

Enrichment plots were generated using GSEA software using the list of genes intersected by a Tloβ2 peaks as the input (ranked by peak score). Plots depict the enrichment score (top, green) for the genes across the ranked list, with the black bars indicating where each gene in the set falls in the list. (A) Enrichment plot for "CA21CHR3X1" the gene set, between positions 1-250,000 bp of chromosome 3. (B) Enrichment plot for the "CELLULAR CARBOHYDRATE CATABOLIC PROCESS_BIO" gene set (GO:0044275). (C) Enrichment plot for the "INDUCTION BY SYMBIONT OF HOST DEFENSE RESPONSE_BIO" gene set (GO:0044416). (D) Enrichment plot for the "YEAST-FORM CELL WALL" gene set (GO:0030445).

DEFENSE RESPONSE_BIO" (genes in GO:0044416), Figure 7.15 (C). Genes which Tloβ2 intersected were also enriched for gene sets that contained genes involved in the structure of the cell wall such as "YEAST-FORM CELL WALL_CEL" (genes in GO:0030445), Figure 7.15 (D), "RAS1_YEAST_UP" which included genes upregulated in *ΔΔras1* yeast cells, and "YAK1_Y_UP" containing genes upregulated in *ΔΔyak1* yeast cells. There was also significant enrichment of the "Gal4_TF" group, which contains genes bound by the Gal4 transcription factor, and the "MED31_UP" set, which contains genes upregulated in a *ΔΔmed31* mutant. The "MED31_DN" set was also enriched but was not statistically significant.

In the case of genes that Tloγ11 interacted with, 2,015 gene sets were enriched, with 157 being statistically significant. There was enrichment of many gene sets at the ends of chromosomes, with interactions at the ends of all chromosomes except chromosome R. Some of the most significantly enriched sets were at the ends of chromosomes 1, 3 and 4, for example "CA21CHR4X9", "CA21CHR1X17", Figure 7.16 (A) and "CA21CHR3X1". There was also enrichment of gene sets related to the negative regulation of cellular processes, such as "NEGATIVE REGULATION OF TRANSCRIPTION FROM RNA POLYMERASE II PROMOTER_BIO" (genes in GO:0000122) and "NEGATIVE REGUALTION OF CELLULAR METABOLIC PROCESS_BIO" (genes in GO:0031324), Figure 7.16 (B). Gene sets related to cellular adhesion were also enriched, such as "CELL ADHESION_BIO" (genes in GO:0007155), Figure 7.16 (C) and "BIOLOGICAL ADHSEION_BIO" (genes in GO:0022610). The set of genes bound by the Efg1 transcription factor were also found to be enriched, "EFG1_TF", Figure 7.16 (D), as was the gene set "EFG1_HYPHAE_DN" which are genes downregulated in hyphal cells of an *ΔΔefg1* mutant.

7.3.4.4 Relationship between Tlo interaction and gene expression

Raw expression values of the genes intersected by peaks from each Tlo protein in each strain were compared to those genes not intersected by the protein peaks. In all cases, genes intersected by Tlo protein peaks were expressed at a higher level than those that were not intersected, Figure 7.17.

Genes intersected by peaks from each Tlo were also compared to the differential expression analysis from Chapter 6. Lists of significantly up- or downregulated genes in each strain compared to the *ΔΔtlo* mutant were analysed to determine if Tlo binding had an influence in up or down regulation of these genes. In the case of Tloα1, 6% of genes

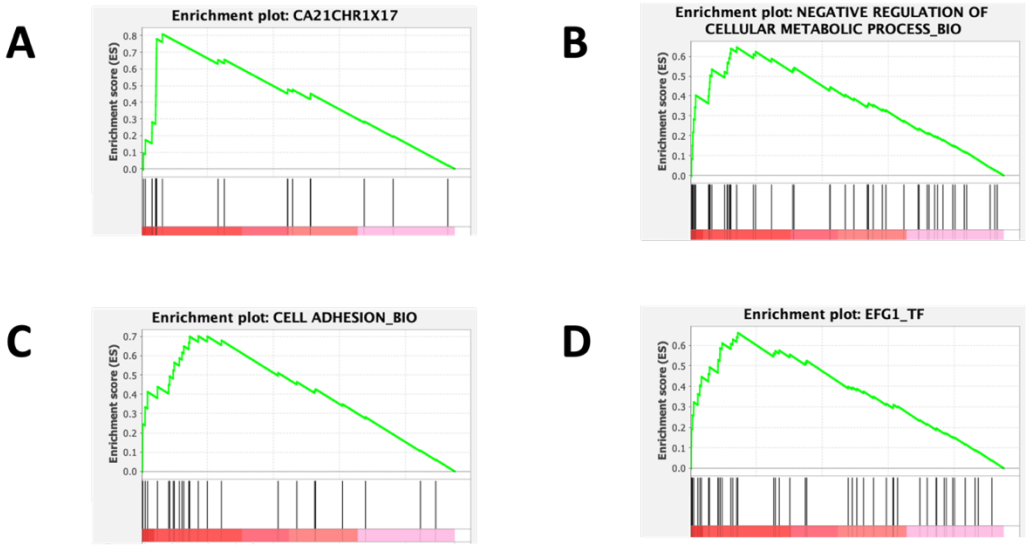


Figure 7.16 Enrichment plots of gene sets enriched in genes which Tloγ11 peaks intersected

Enrichment plots were generated using GSEA software using the list of genes intersected by a Tloγ11 peaks as the input (ranked by peak score). Plots depict the enrichment score (top, green) for the genes across the ranked list, with the black bars indicating where each gene in the set falls in the list. (A) Enrichment plot for the "CA21CHR1X17" gene set, between positions 3,150,000-3,450,000 bp of chromosome 1. (B) Enrichment plot for the "NEGATIVE REGUALTION OF CELLULAR METABOLIC PROCESS_BIO" gene set (GO:0031324). (C) Enrichment plot for the "CELL ADHESION_BIO" gene set (GO:0007155). (D) Enrichment plot for the "EFG1_TF" gene set, genes bound by the Efg1 transcription factor.

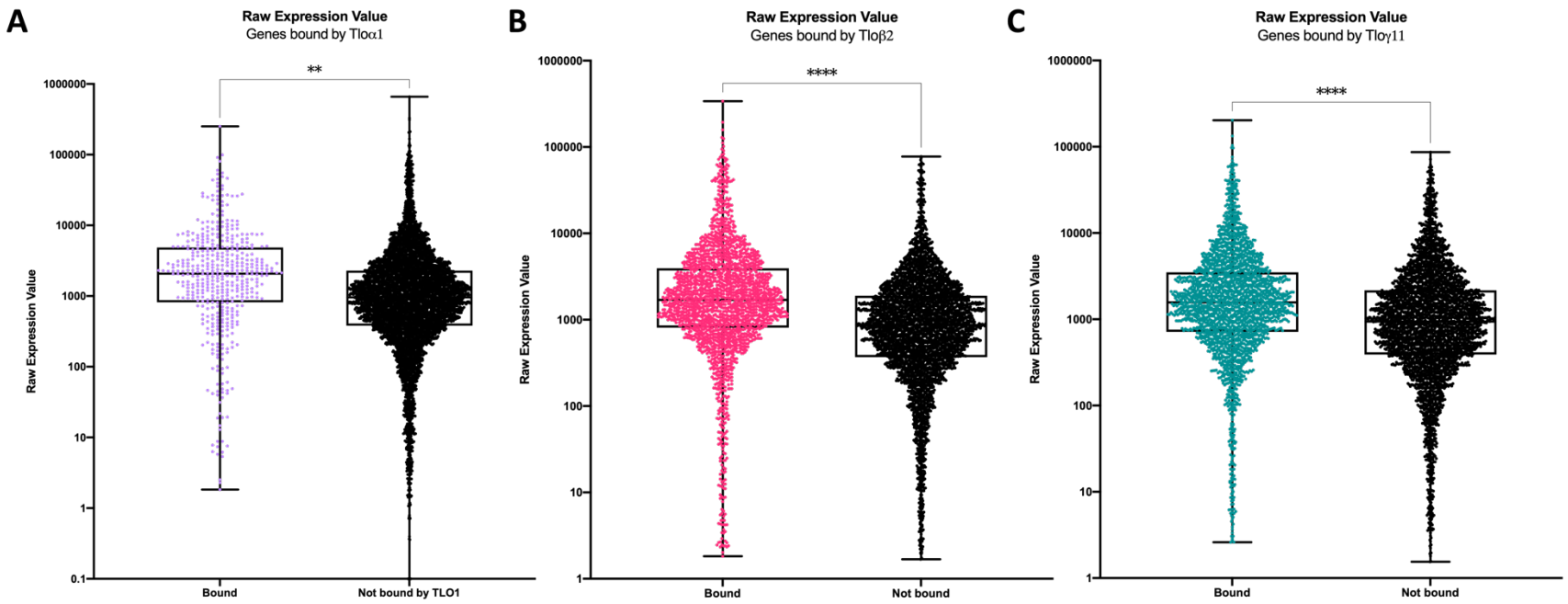


Figure 7.17 Expression of genes intersected and those not intersected by Tlo protein peaks in the *TLO* reintroduction strains

The raw expression level from the RNA-sequencing data for each gene identified as being intersected by a Tlo protein peak in each *TLO* reintroduction strain was graphed, as well as the raw expression value for all other genes (not intersected) in each strain. Intersected genes were either intersected at the ORF, at the TSS or at the promoter. Statistical significance was calculated using an unpaired *t* test with Welch's correction, bars with asterisks denote statistically significant difference between groups. (A) Raw expression values of genes intersected by Tloα1 peaks, and those not intersected, in the *ΔΔtlo::P_{TET}TLOα1* strain. (B) Raw expression values of genes intersected by Tloβ2 peaks, and those not intersected, in the *ΔΔtlo::P_{TET}TLOβ2* strain. (C) Raw expression values of genes intersected by Tloγ11 peaks, and those not intersected, in the *ΔΔtlo::P_{TET}TLOγ11* strain.

that were upregulated in the *ΔΔtlo::P_{TET}TLOα1* strain compared to the *ΔΔtlo* mutant were intersected by Tloα1 protein peaks, while 8% of downregulated genes were intersected. These significantly altered genes accounted for 33% of all of the Tloα1 protein's interactions with genes. For Tloβ2, 24% of genes upregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain compared to the *ΔΔtlo* mutant, were intersected by Tloβ2 protein peaks, while 41% of downregulated genes were intersected by Tloβ2 protein peaks. Here, significantly up or downregulated genes accounted for 19% of Tloβ2 protein-gene interactions. For Tloγ11, 66% of genes that were upregulated in the *ΔΔtlo::P_{TET}TLOγ11* strain compared to the *ΔΔtlo* mutant were intersected by Tloγ11 protein peaks, and 55% of downregulated genes were intersected. In this case, significantly altered genes accounted for >1% of the Tloγ11 gene interactions.

There were 48 genes that were uniquely intersected by Tloα1 peaks and upregulated in the *ΔΔtlo::P_{TET}TLOα1* strain, Figure 7.18 (A), however, no GO enrichment was found within this list. There were 124 genes that were uniquely intersected by Tloβ2 peaks that were upregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain, Figure 7.18 (A). These genes were enriched for GO terms related to carbohydrate metabolism and small molecule biosynthesis, as well as GO terms related to the defence response (GO:0031347), with genes *ENO1, RIM101, CDC19, IMH3, TDH3* and *FBA1* falling under these terms, Figure 7.19. Of the nine genes that were uniquely intersected by Tloγ11 peaks and upregulated in the *ΔΔtlo::P_{TET}TLOγ11* strain, only the GO term for anion transmembrane transport (GO:0098656) was enriched, with the *CTP1* and *GIT3* genes falling under this term, Figure 7.18 (A).

Tloα1 and Tloβ2 bind 19 common genes that were upregulated in each respective strain, Figure 7.18 (A). This list is enriched for DNA-binding transcription factors, specifically *CRZ1, CUP9, FCR1, MIG1* and *TYE7*. There were seven genes that were intersected by both Tloβ2 and Tloγ11 peaks and upregulated in their respective strains, with significant enrichment for the cellular component GO term cell surface (GO:0009986) with *FET99, PGA23, PGA57, PHR2* falling under this term. Tloα1 and Tloγ11 did not share genes that were intersected and upregulated.

Tloα1 uniquely interacted with 67 genes that were significantly downregulated in the *ΔΔtlo::P_{TET}TLOα1* strain, Figure 7.18 (B). These genes were subjected to GO analysis, with enrichment of GO terms related to carbohydrate transport (GO:0008643), localisation (GO:0051179) and autophagy (GO:0006914) found, Figure 7.20 (A).

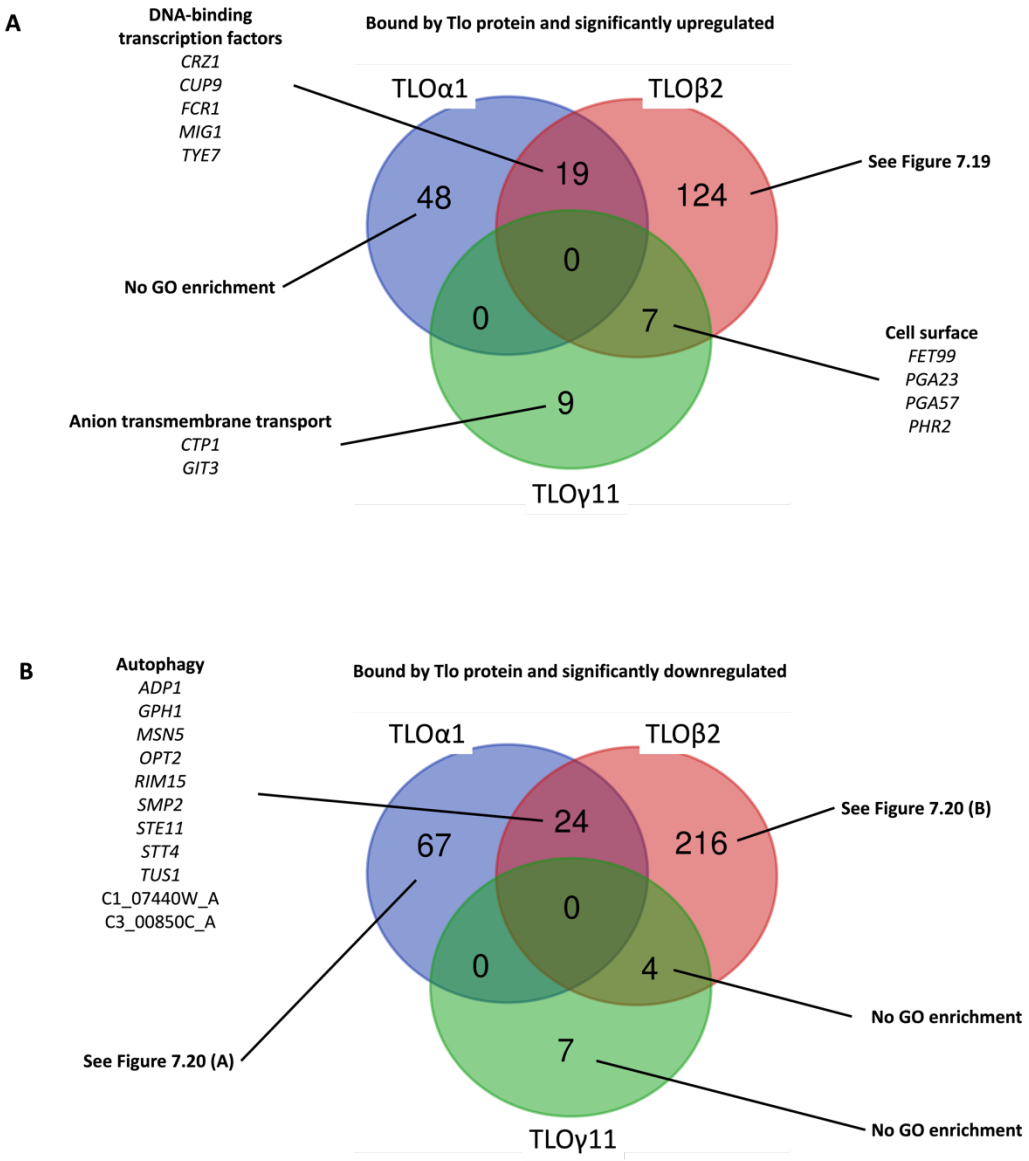


Figure 7.18 Genes interacted with and putatively regulated by Tlo proteins

Genes that the Tlo proteins interacted with and which were either significantly upregulated or downregulated in the relative strains when compared to the *ΔΔtlo* mutant (from RNA-seq data) were compiled into Venn diagrams. Distinct sets were subjected to GO analysis, and significant terms and genes included in them were added to the diagrams (A) Genes Tlo proteins interacted with that were upregulated in their respective strains. (B) Genes that Tlo proteins interacted with and were downregulated in their respective strains.

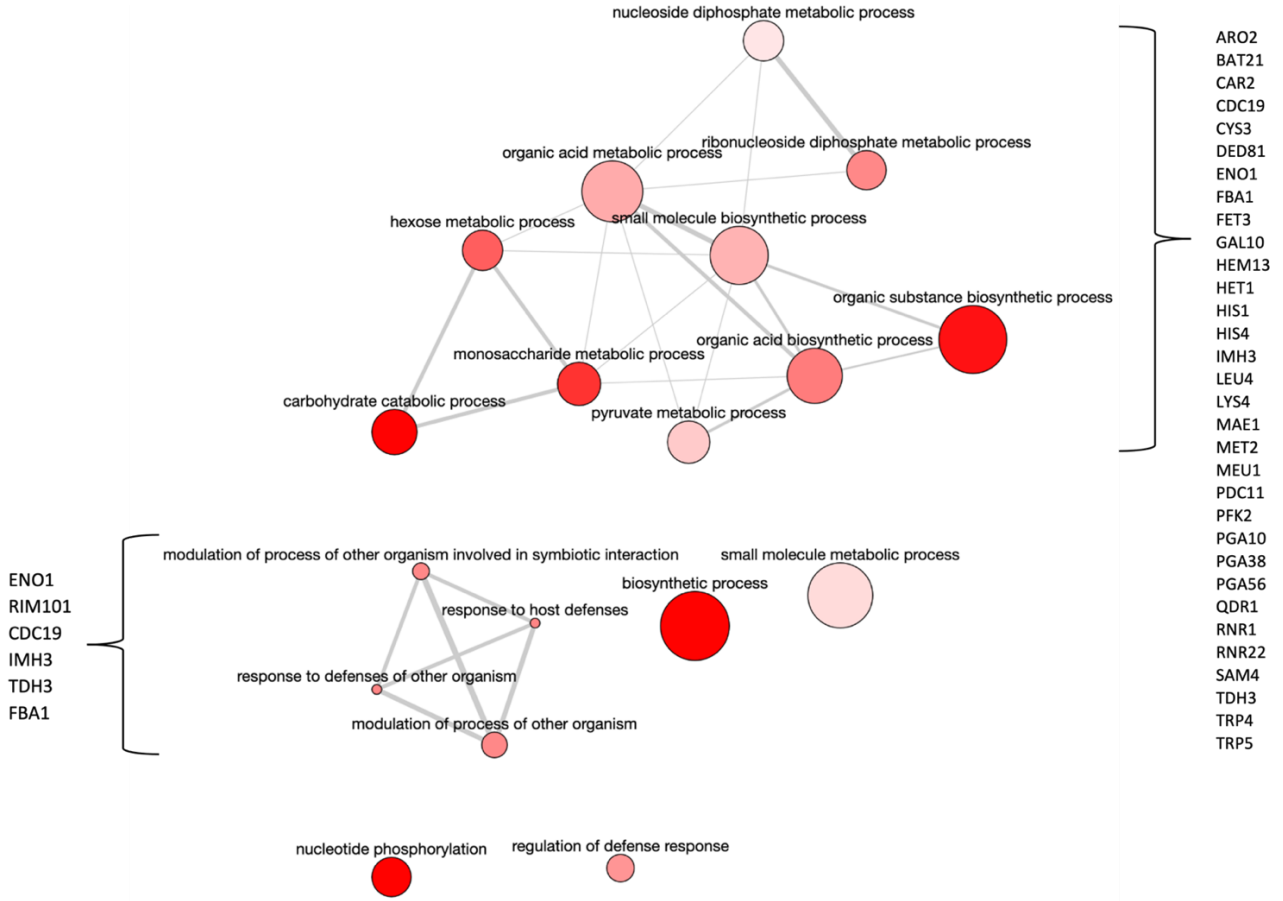


Figure 7.19 Genes which Tloβ2 interacted with which were upregulated in the *ΔΔtlo::P$_{TET}$TLOβ2* strain (cont.)

Genes which Tloβ2interacted with which were upregulated in the *ΔΔtlo::P$_{TET}$TLOβ2* strain compared to the *ΔΔtlo* mutant were subjected to GO analysis. Enriched biological process GO terms are displayed. Many of the GO terms in the same cluster contain similar genes, and these genes have been listed next to each relative cluster.

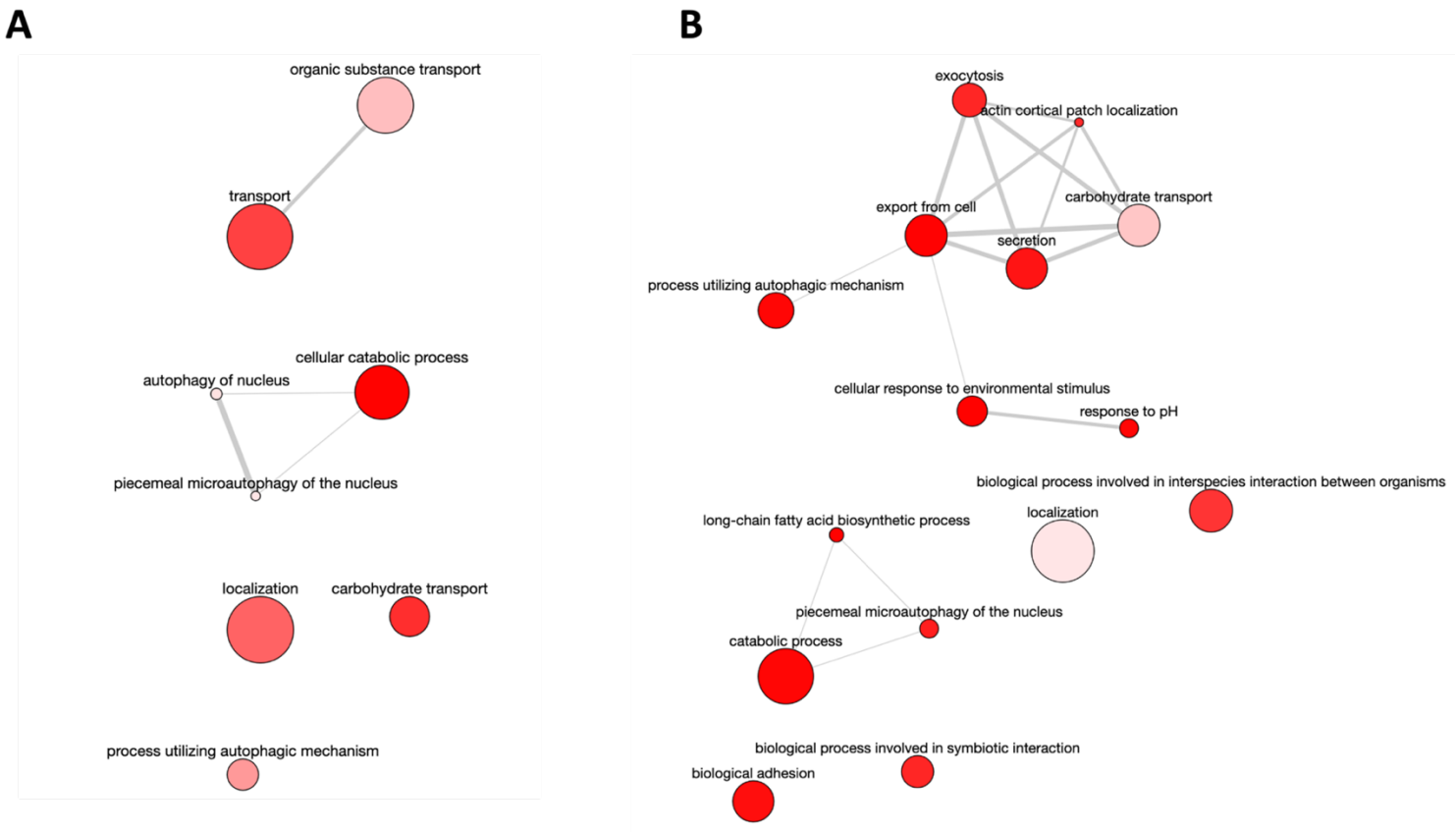


Figure 7.20 Genes which Tloα1 or Tloβ2 interacted with which were downregulated in respective strains (cont.)

Genes which Tloα1 or Tloβ2 interacted with which were downregulated in the respective strain compared to the *ΔΔtlo* mutant were subjected to GO analysis. Enriched biological process GO terms are displayed. (A) GO analysis of genes which Tloα1 interacted with which were downregulated in the *ΔΔtlo::P_{TET}TLOα1* strain. (B) GO analysis of genes which Tloβ2 interacted with which were downregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain.

Genes *AHR1, ARG1, CYB2, DEM1, FMO1, HGT1, IFF5, PDK2, PEX6, PGA49, PNG2, PXA1, PXA2*, *SAP6* and *TRY4* contributed to the enrichment of the carbohydrate transport term, with many also enriching the localisation term. Several autophagy terms were enriched, including autophagy of the nucleus (GO:0044804) and autophagy of the mitochondrion (GO:0000422). Genes *BPH1, CIRT4B, CYR1, IFF5, KOG1, MHP1, PEP1, PEP3, PEX6, PTR3, PXA1, PXA2, SCT2* and *VAM6* enriched these autophagy GO terms.

Tloβ2 uniquely interacted with 216 genes that were significantly downregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain, Figure 7.18 (B). These genes were enriched for GO terms related to autophagy (GO:0006914) and carbohydrate transport (GO:0008643), as well as several terms related to the cellular response to external stimuli (e.g. cellular response to environmental stimulus (GO:0104004)), Figure 7.20 (B). Genes *HGT8, HGT12, HGT14, HGT18, HGT19, HXK1* and *HXT5* contribute to the enrichment of the carbohydrate transport term and genes *ALS1, BEM3, FAS2, HGT12, HSP104, MDS3, NIK1, PAN1, PHR1, RHA1, RIM13, SLK19, SOG2* and *SSY5* contributing to the response to environmental stimulus term.

For Tloγ11, 7 genes were uniquely intersected by protein peaks and downregulated in the *ΔΔtlo::P_{TET}TLOγ11* strain, *CTR1* and *ENA2* and five uncharacterised ORFs, Figure 7.18 (B), however there was no GO enrichment found in this list.

Tloα1 and Tloβ2 bind 24 common genes that were downregulated in both respective strains, Figure 7.18 (B), and analysis of this list shows enrichment for GO terms related to autophagy (GO:0006914) with *ADP1, GPH1, MSN5, OPT2, RIM15, SMP2, STE11, STT4, TUS1* and C1_07440W_A and C3_00850C_A falling under this term. Many of these genes also fall under the GO terms for transport (GO:0006810). There were four genes that were intersected by both Tloβ2 and Tloγ11 peaks that were downregulated in their respective strains, *RGA2*, *ZCF2* and uncharacterised ORFs C7_01510W_A and CR_07140C_A, however this list did not show enrichment for any GO terms. Again, the Tloα1 and Tloγ11 proteins both did not bind any genes that were commonly downregulated in these strains.

The lists of gene with significantly altered gene expression which the Tlo proteins interacted with were submitted to GSEA, ranked by expression value in the relative *TLO* reintroduction strain compared to the *ΔΔtlo* mutant, from Chapter 6. There was no enrichment found in the dataset intersected by Tloγ11 peaks.

With regards to genes that Tloα1 interacted with, there were 59 gene sets enriched in the upregulated genes, with 19 being significant, while there were 39 gene sets enriched in the downregulated gene set, with two of these being significant. The most significantly enriched gene sets in the upregulated dataset included "WHITE_UP", "XS_UP", "TYE7_TF", and "MED31_DN", Figure 7.21 (A-D), as well as gene sets bound by transcription factors such as "BRG1_TF". The two gene sets that were significantly enriched in the downregulated dataset were the "KETOCONAZOLE_UP" and the "RHE90_UP" gene sets, Figure 7.21 (E&F), the "UPC2_UP" gene set is also enriched, but not significantly.

Genes that were intersected by Tloβ2 peaks which were upregulated were found to be enriched for 105 gene sets, with 40 of these being significant, while genes which were downregulated were enriched for 136 gene sets, with 33 being significant. The most significantly enriched gene sets in the upregulated gene dataset included "KETOCONAZOLE_DN", "HYPHAE_DN", "WHITE_UP", "XS_UP", Figure 7.22 (A&B), "BMDM PHAGOCYTOSIS_DN", "TYE7_TF" and "MED31_DN", Figure 7.22 (C&D). Those that were enriched in the downregulated dataset included "KETOCONAZOLE_UP" and "RHE90_UP", Figure 7.22 (E&F), "OPAQUE_UP", and "BMDM PHAGOCYTOSIS_UP", along with many sets related to the hyphal morphology such as "HYPHAE_SPIDER_UP", "HYPHAE_UP" and "HYPHAL GROWTH_BIO" (genes in GO:0030448).

Genes that were intersected by either Tloα1 or Tloβ2 peaks, which were significantly differentially expressed compared to the *ΔΔtlo* mutant., in groups that were highlighted by GSEA as being significantly enriched, were analysed and their expression values examined. The expression level of genes in the gene sets WHITE_UP", "XS(2006)_UP", and "MED31_DN", that were intersected by either Tloα1, Tloβ2 peaks or both, were graphed, and heatmaps were generated for the expression levels of individual genes, Figure 7.23 (A-F). These gene sets were intersected by the Tlo protein peaks and were significantly upregulated in their respective strains.

In the "WHITE_UP" data set, genes intersected by Tloα1 peaks were generally expressed at a higher level than those intersected by Tloβ2 peaks, Figure 7.23 (A). All genes that Tloα1 interacted with were upregulated, while a subset of genes that Tloβ2 interacted with were downregulated, including *AAF1, ASR1&2* and *ADH5,* Figure 7.23 (B). Between the two sets, genes which Tloα1 and Tloβ2 interacted with which were upregulated

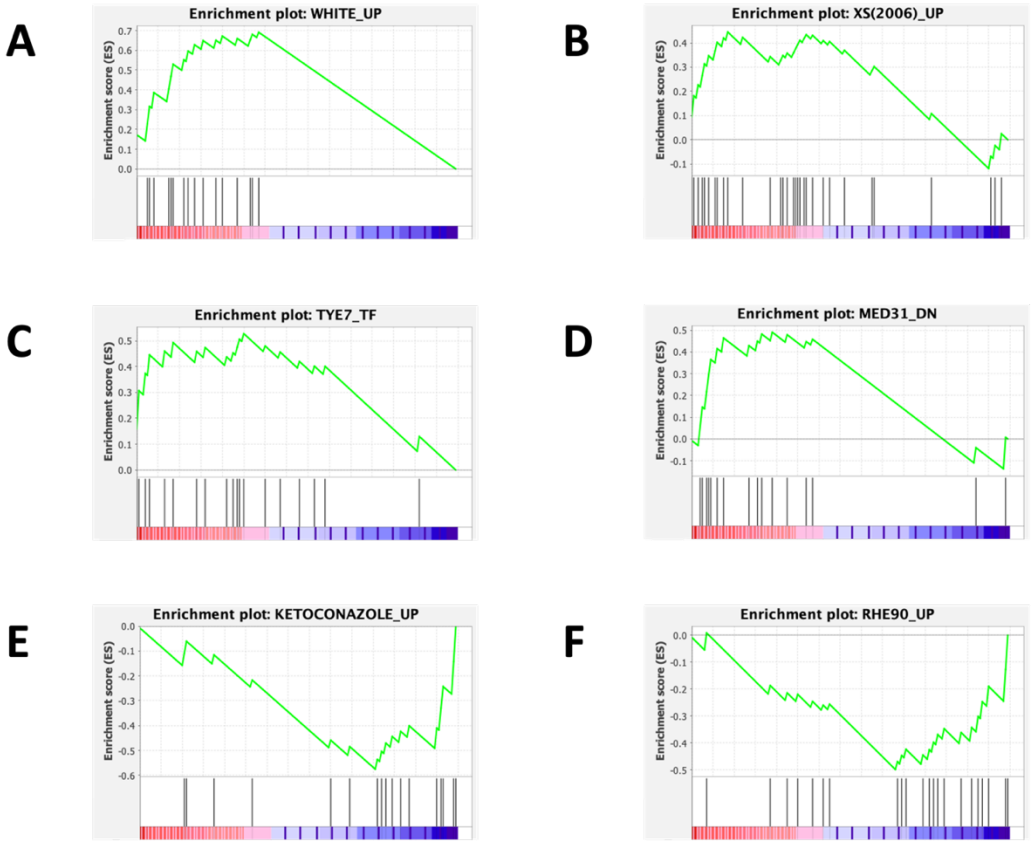


Figure 7.21 Enrichment plots of gene sets transcriptionally altered by Tloα1 interaction

Enrichment plots were generated using GSEA software using the list of genes intersected by a Tloα1 peaks which were also differentially expressed in the *ΔΔtlo::P_{TET}TLOα1* strain compared to the *ΔΔtlo* mutant as the input (ranked by L_2FC). Plots depict the enrichment score (top, green) for the genes across the ranked list, with the black bars indicating where each gene in the set falls in the list. Positive ES indicates enrichment in the upregulated dataset, and negative indicates enrichment in the downregulated dataset. (A) Enrichment plot for the "WHITE_UP" gene set, genes upregulated in white cells. (B) Enrichment plot for the "XS(2006)_UP" gene set, genes upregulated upon exposure to oxidative stress. (C) Enrichment plot for the "TYE7_TF" gene set, genes bound by the Tye7 transcription factor. (D) Enrichment plot for the "MED31_DN" gene set, genes downregulated in a *ΔΔmed31* mutant. (E) Enrichment plot for the "KETOCONAZOLE_UP" gene set, genes upregulated upon exposure to ketoconazole. (F) Enrichment plot for the "RHE90_UP" gene set, genes upregulated after 90 mins exposure to reconstituted human epithelium.

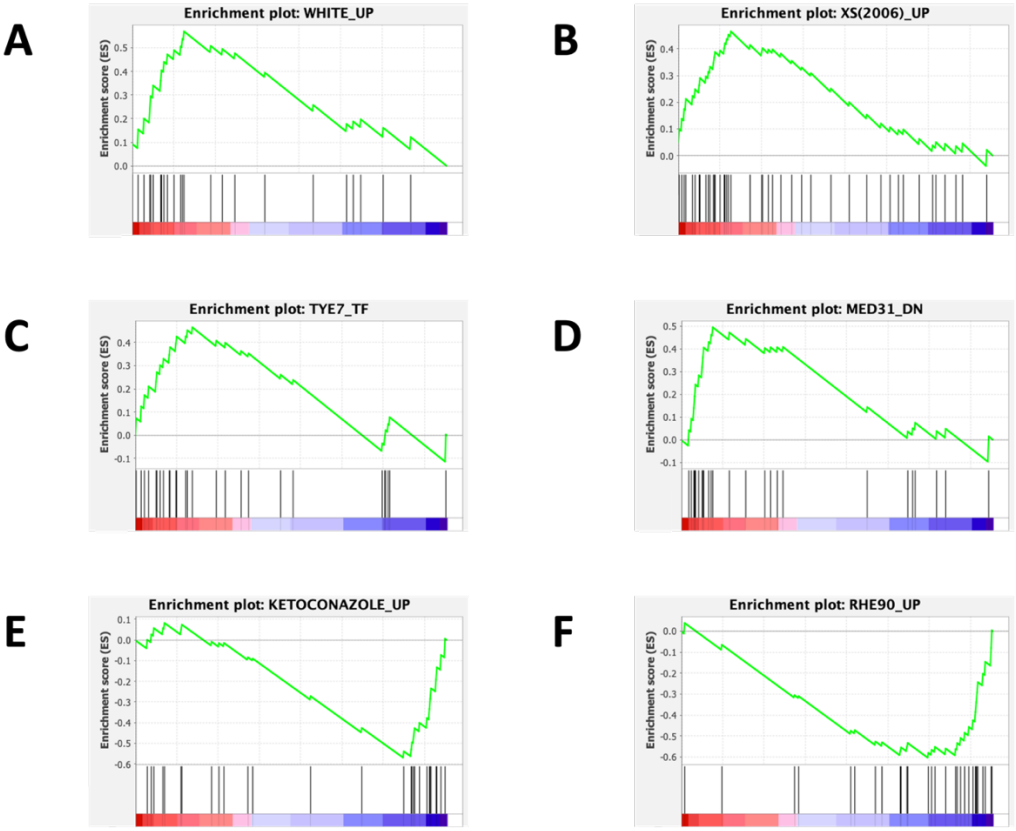


Figure 7.22 Enrichment plots of gene sets transcriptionally altered by Tloβ2 interaction

Enrichment plots were generated using GSEA software using the list of genes intersected by a Tloβ2 peaks which were also differentially expressed in the *ΔΔtlo::P$_{TET}$TLOβ2* strain compared to the *ΔΔtlo* mutant as the input (ranked by L_2FC). Plots depict the enrichment score (top, green) for the genes across the ranked list, with the black bars indicating where each gene in the set falls in the list. Positive ES indicates enrichment in the upregulated dataset, and negative indicates enrichment in the downregulated dataset. (A) Enrichment plot for the "WHITE_UP" gene set, genes upregulated in white cells. (B) Enrichment plot for the "XS(2006)_UP" gene set, genes upregulated upon exposure to oxidative stress. (C) Enrichment plot for the "TYE7_TF" gene set, genes bound by the Tye7 transcription factor. (D) Enrichment plot for the "MED31_DN" gene set, genes downregulated in a *ΔΔmed31* mutant. (E) Enrichment plot for the "KETOCONAZOLE_UP" gene set, genes upregulated upon exposure to ketoconazole. (F) Enrichment plot for the "RHE90_UP" gene set, genes upregulated after 90 mins exposure to reconstituted human epithelium.

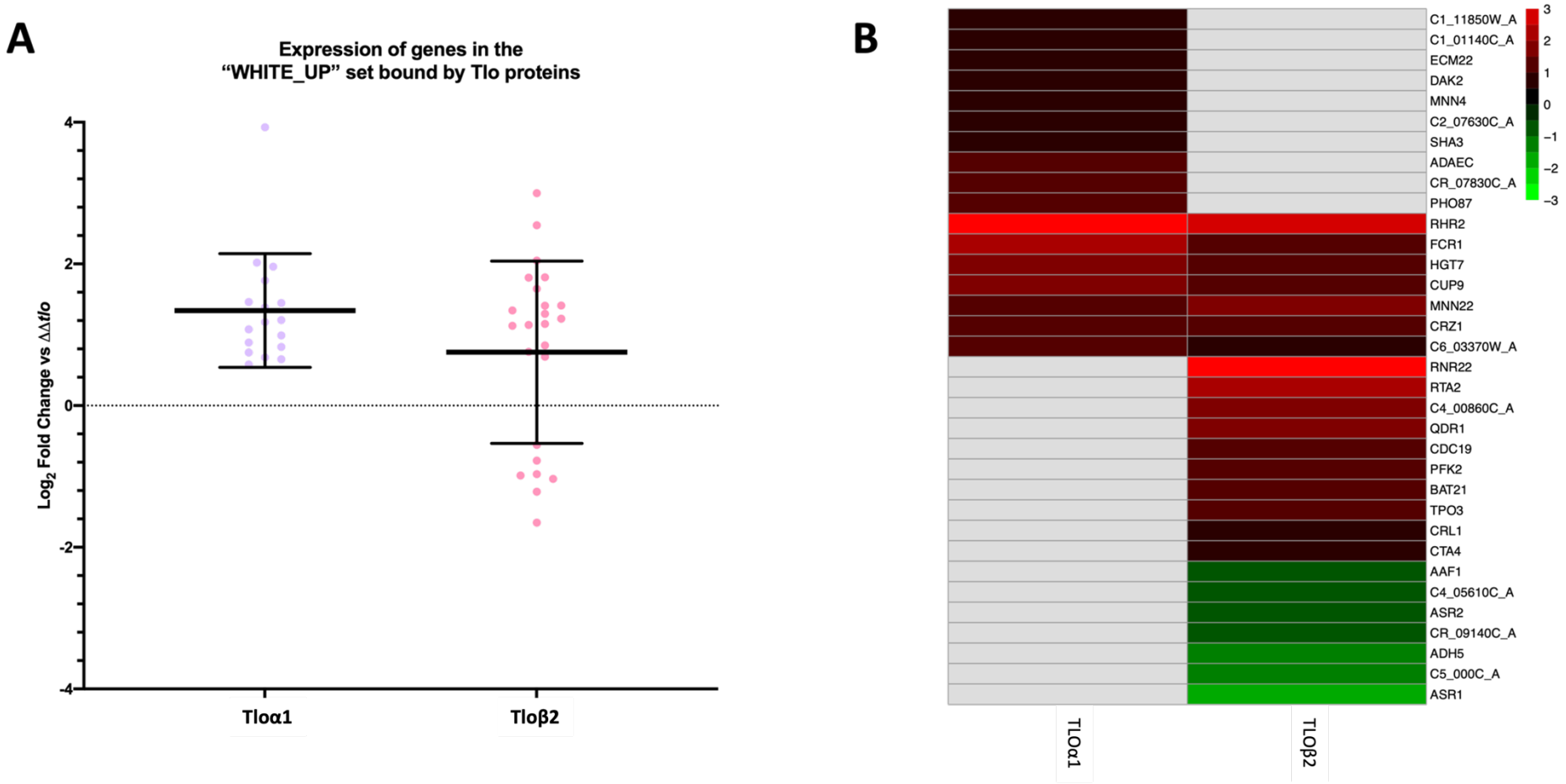


Figure 7.23 Expression of genes which Tloα1 or Tloβ2 interact with in upregulated gene sets (cont.)

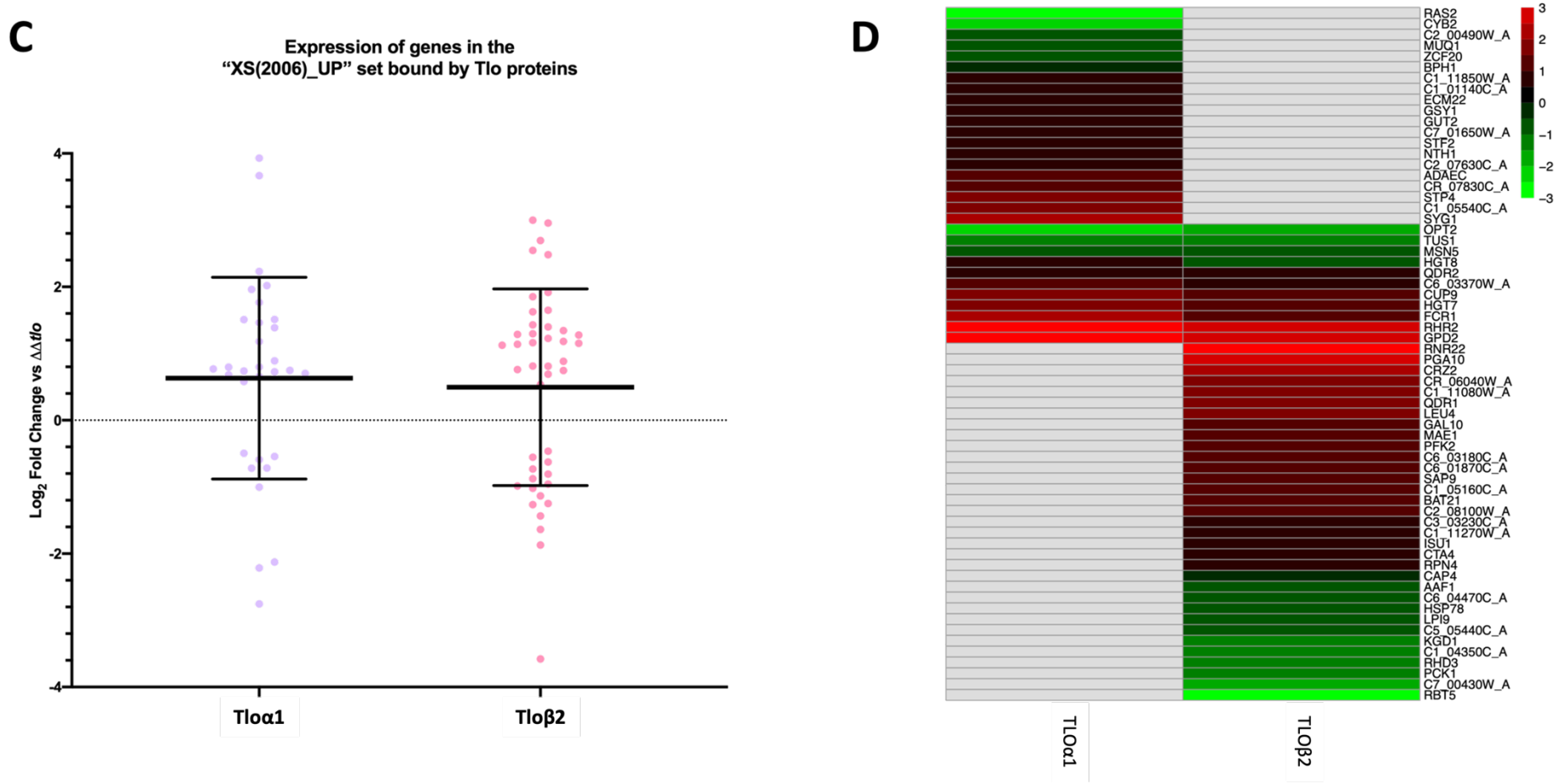


Figure 7.23 Expression of genes which Tloα1 or Tloβ2 interact with in upregulated gene sets (cont.)

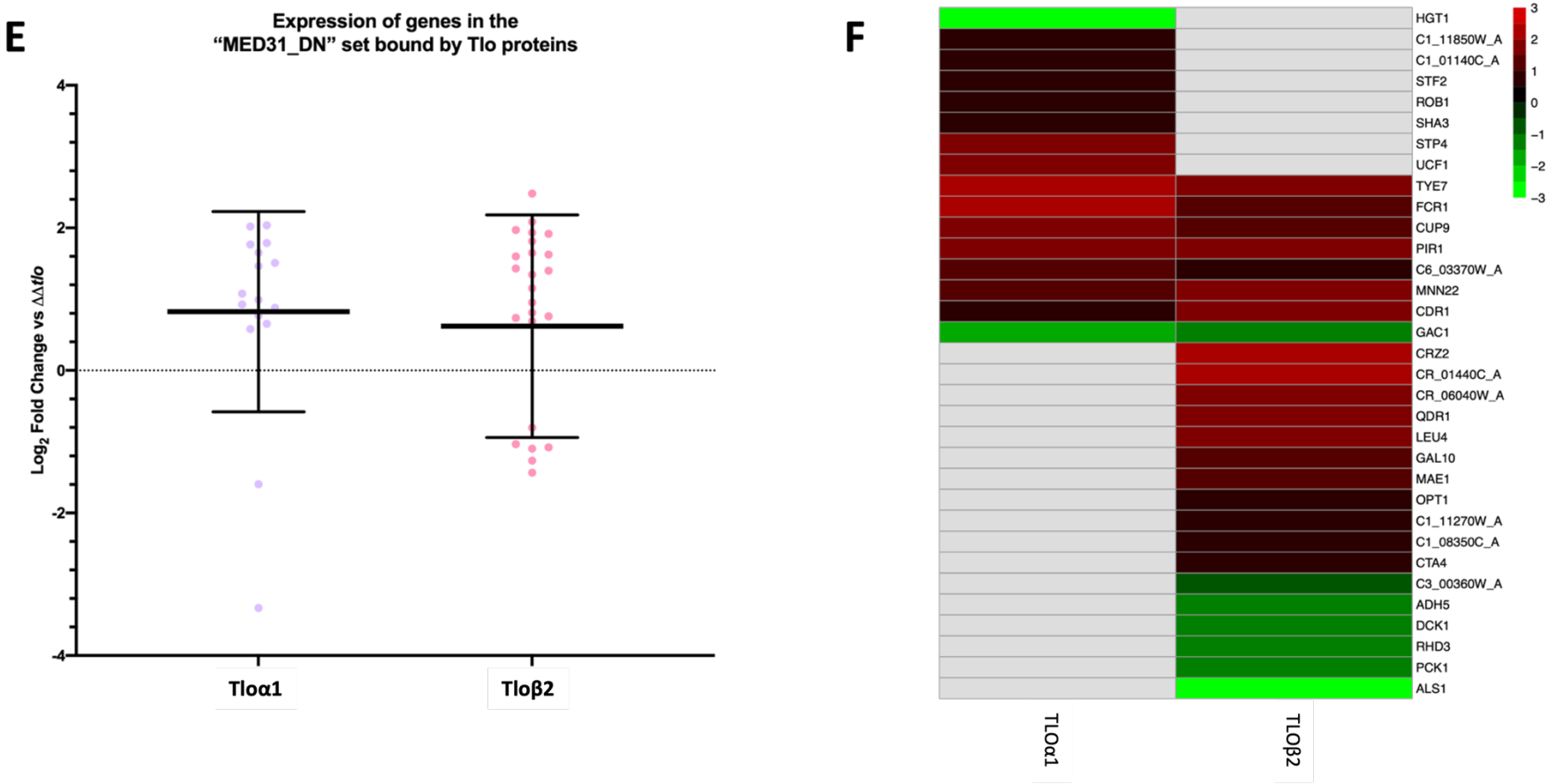


Figure 7.23 Expression of genes which Tloα1 or Tloβ2 interact with in upregulated gene sets (cont.)

Figure 7.23 Expression of genes which Tloα1 or Tloβ2 interact with in upregulated gene sets (cont.)

The L_2FC of genes that were intersected by either Tloα1, Tloβ2 peaks or both, which were significantly differentially expressed compared to the *ΔΔtlo* mutant., in groups that were highlighted by GSEA as being significantly enriched (upregulated) when intersected genes were analysed, are plotted, unpaired *t* tests with Welsh's correction were performed to determine if difference in the mean expression level were significant (none found). Expression values for individual genes are shown in heatmaps. Heatmaps indicate the L_2FC compared to the *ΔΔtlo* mutant, with genes upregulated in red and downregulated in green (see scale on right), grey indicates that either the gene was not intersected, or was not significantly differentially expressed. (A) Expression of genes in the "WHITE_UP" gene set intersected by Tloα1 and Tloβ2 peaks. These are genes typically upregulated in white cells. (B) Heatmap of expression of genes in the "WHITE_UP" gene set intersected by Tloα1 and Tloβ2 peaks. (C) Expression of genes in the "XS(2006)_UP" gene set intersected by Tloα1 and Tloβ2 peaks. These are genes upregulated after exposure to oxidative stress. (D) Heatmap of expression of genes in the "XS(2006)_UP" gene set intersected by Tloα1 and Tloβ2 peaks. (E) Expression of genes in the "MED31_DN" gene set intersected by Tloα1 and Tloβ2 peaks. These are genes downregulated in a *ΔΔmed31* mutant (F) Heatmap of expression of genes in the "MED31_DN" gene set intersected by Tloα1 and Tloβ2 peaks.

included the transcriptional regulators *FCR1, CUP9* and *CRZ1,* Figure 7.23 (B). In the "XS(2006)_UP" data set, many of the genes intersected by Tloα1 and Tloβ2 peaks in this set were downregulated as well as upregulated, but the mean fold change of these intersected genes compared to the *ΔΔtlo* was indicative of upregulation, Figure 7.23 (C). Expression of genes that were intersected by peaks from both Tlos were found to be similar between the two strains, i.e. genes intersected by Tloα1 peaks which were found to be upregulated, were also upregulated and intersected by Tloβ2 peaks, with these genes including transcriptional regulators *FCR1* and *CUP9*, Figure 7.23 (D). These genes made up only a small subset of this gene set, and overall, the oxidative stress "UP" gene sets were enriched in the *ΔΔtlo* when the entire transcriptome was accounted for, see Chapter 6.

The genes in the "MED31_DN" set that the Tlos interacted with were also generally upregulated, Figure 7.23 (E), although there were subsets of genes that were downregulated. Both Tlos interacted with the *GAC1* gene, and this gene was downregulated in both strains compared to the *ΔΔtlo* mutant. Genes that only Tloβ2 interacted with which were downregulated included the genes *ADH5, ALS1, DCK1, PCK1* and *RHD3*. Genes that were upregulated which both Tlos interacted with again included transcriptional regulators such as *TYE7, FCR1* and *CUP9*, Figure 7.23 (F).

This analysis was performed on the genes in the "TYE7_TF" and "GAL4_TF" gene sets, Figure 7.24 (A-D). The "TYE7_TF" gene set was highlighted as being intersected by peaks from both Tlo proteins, and their expression generally upregulated, Figure 7.24 (A). Genes that both Tlos interacted with in this group share similar expression patterns, Figure 7.24 (B). Genes in the "GAL4_TF" gene set which Tloα1 interacted with were generally upregulated, however the mean L_2FC of genes that Tloβ2 interacted with in the set was not as upregulated, and this difference was found to be statistically significant, Figure 7.24 (C). Many genes which Tloβ2 interacted with were downregulated compared to the *ΔΔtlo* mutant, such as *ADH1, FBP1, HSP104, LPD1, MAC1, NCR1, RHD3* and *VID21*, Figure 7.24 (D).

The gene sets "KETOCONAZOLE_UP" and "RHE90_UP "were also subjected to this analysis, as these were gene sets which were highlighted as being intersected by peaks from both Tlos, as well as downregulated in the respective strains compared to the *ΔΔtlo* mutant, Figure 7.25 (A-D). Genes in the "KETOCONAZOLE_UP" group intersected by the Tlo protein peaks were generally downregulated, Figure 7.25 (A), with genes

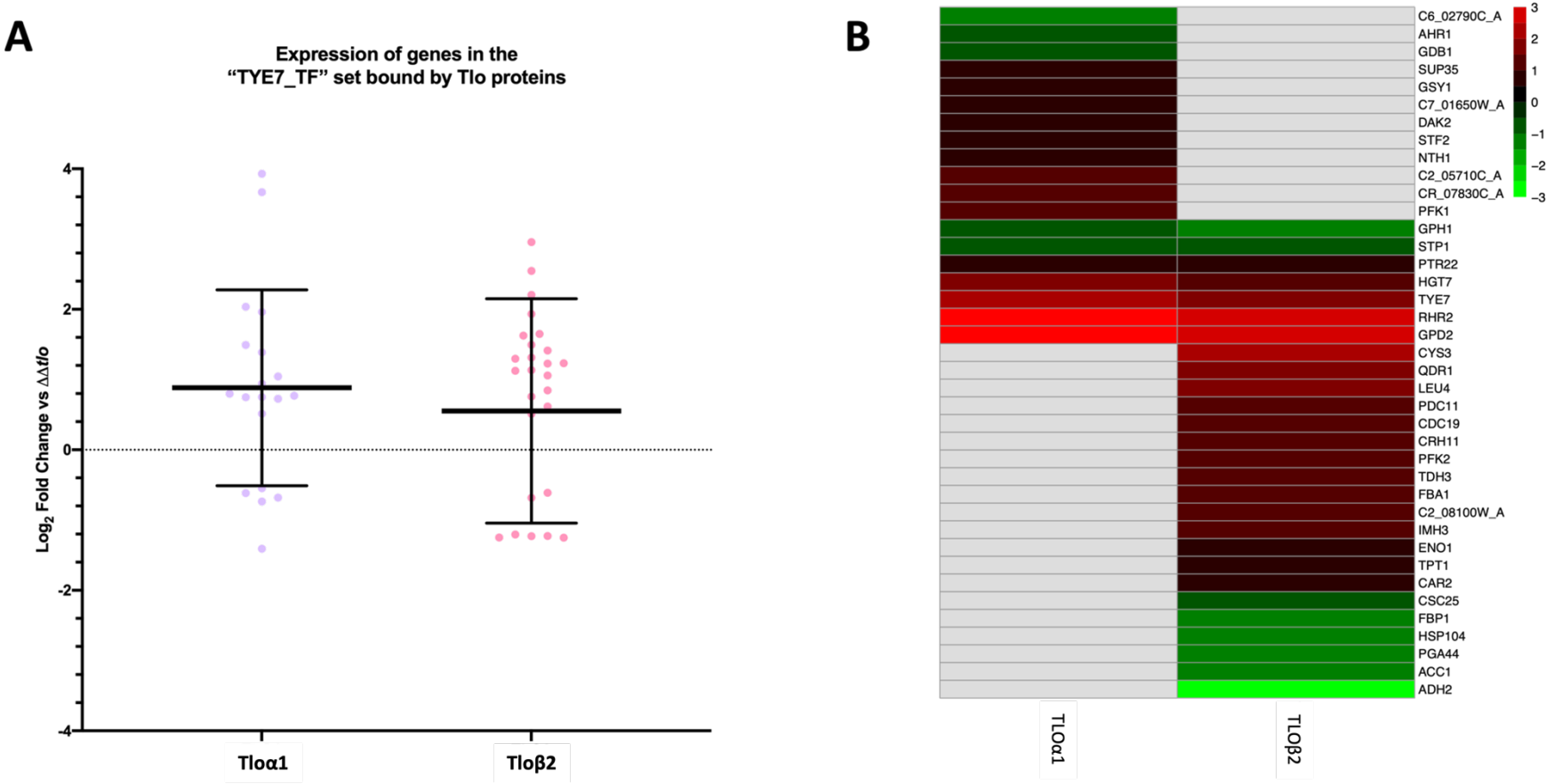


Figure 7.24 Expression of genes which Tloα1 or Tloβ2 interact with in Tye7 and Gal4 bound gene sets (cont.)

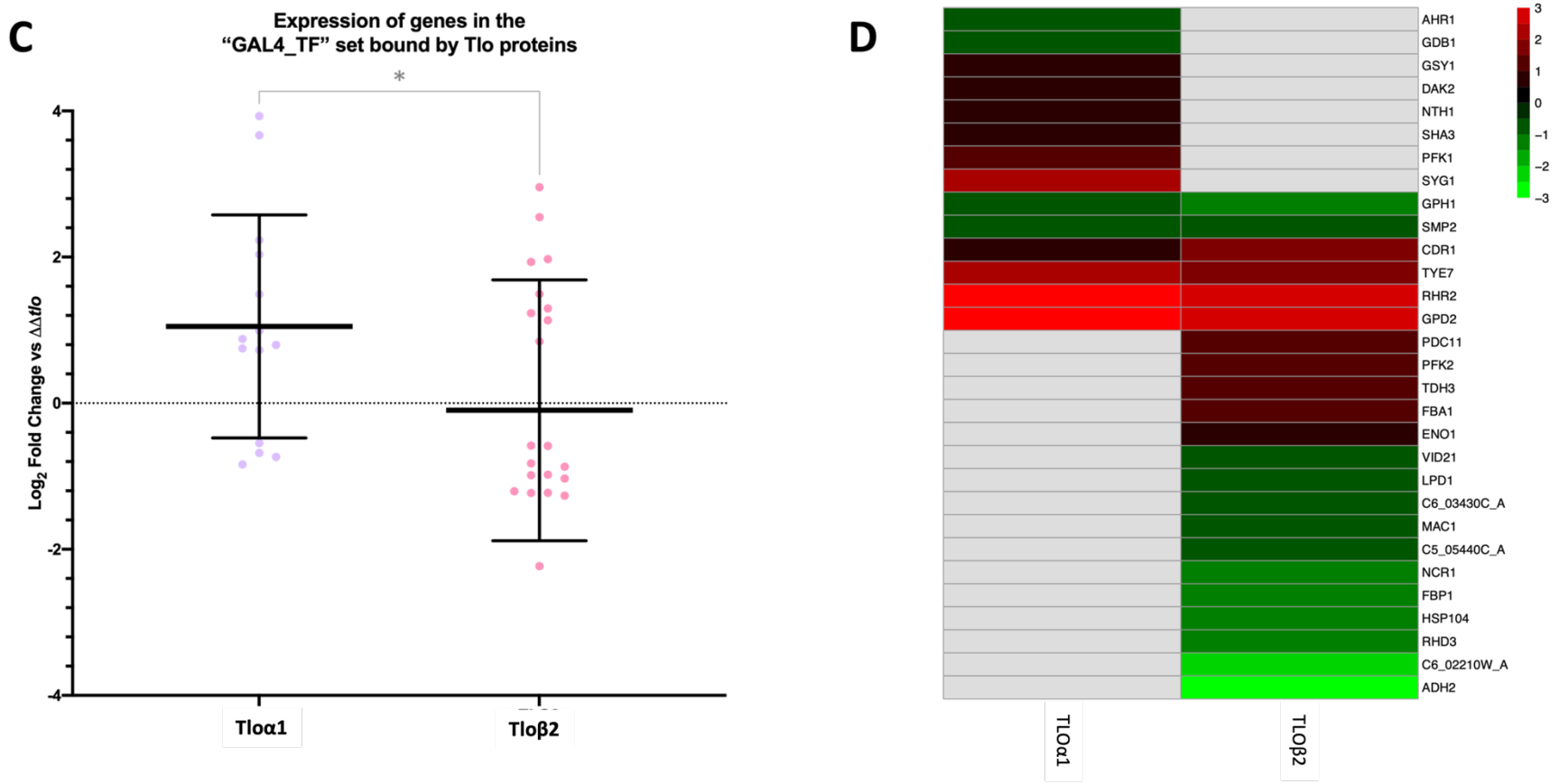


Figure 7.24 Expression of genes which Tloα1 or Tloβ2 interact with in Tye7 and Gal4 bound gene sets (cont.)

Figure 7.24 Expression of genes which Tloα1 or Tloβ2 interact with in Tye7 and Gal4 bound gene sets (cont.)

The L_2FC of genes that were intersected by either Tloα1, Tloβ2 peaks or both, which were significantly differentially expressed compared to the *ΔΔtlo* mutant., in gene sets bound by carbohydrate metabolism regulators Tye7 and Gal4, are plotted, unpaired *t* tests with Welsh's correction were performed to determine if difference in the mean expression level were significant. Expression values for individual genes are shown in heatmaps. Heatmaps indicate the L_2FC compared to the *ΔΔtlo* mutant, with genes upregulated in red and downregulated in green (see scale on right), grey indicates that either the gene was not intersected, or was not significantly differentially expressed. (A) Expression of genes in the "TYE7_TF" gene set intersected by Tloα1 and Tloβ2 peaks. These are genes bound by the Tye7 transcription factor. (B) Heatmap of expression of genes in the "TYE7_TF" gene set intersected by Tloα1 and Tloβ2 peaks. (C) Expression of genes in the "GAL4_TF" gene set intersected by Tloα1 and Tloβ2 peaks. These are genes bound by the Gal4 transcription factor. The difference in expression level between the sets was found to be significant, indicated by an asterisk. (D) Heatmap of expression of genes in the "GAL4_TF" gene set intersected by Tloα1 and Tloβ2 peaks.

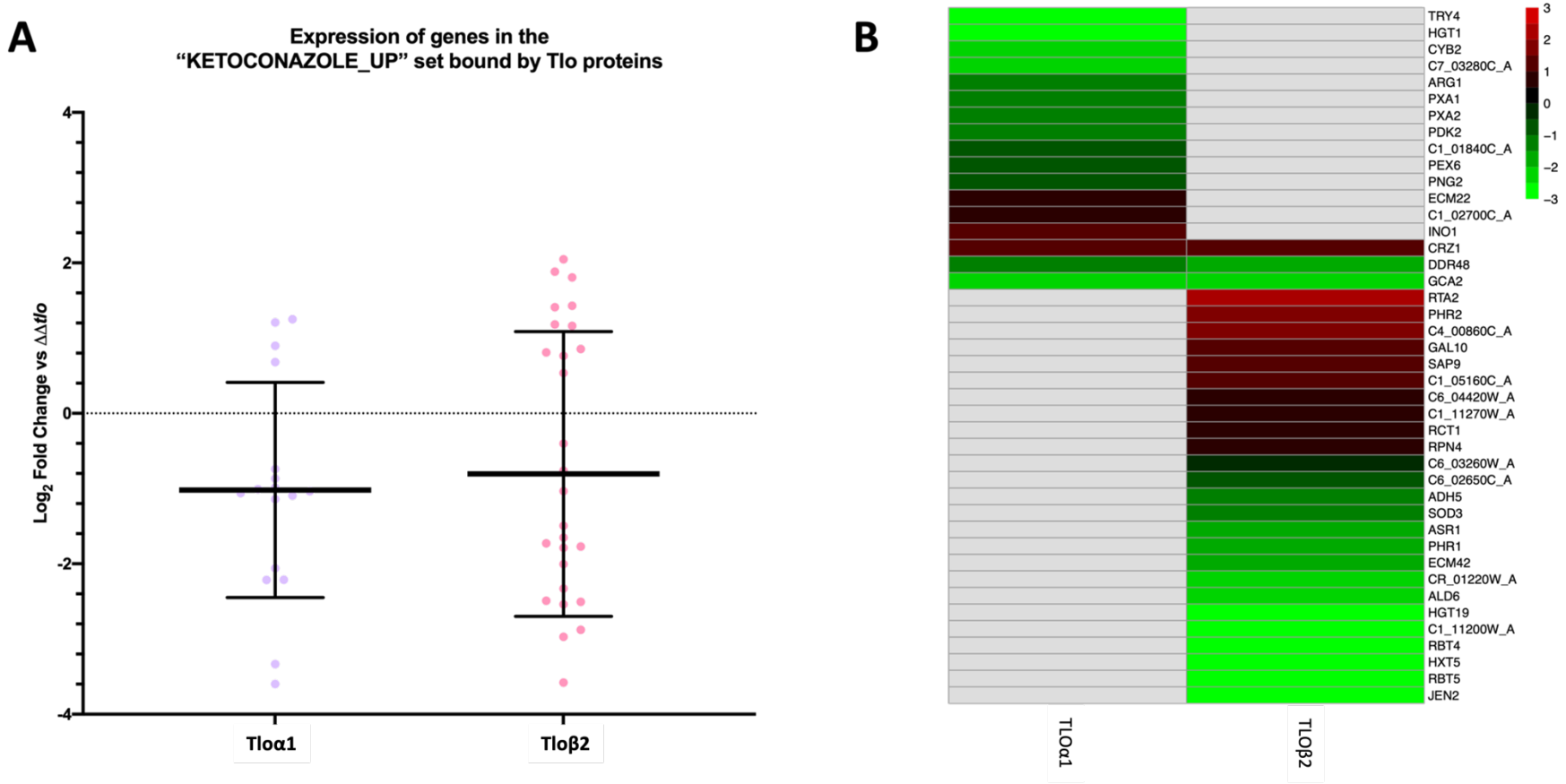


Figure 7.25 Expression of genes which Tloα1 or Tloβ2 interact with in downregulated gene sets (cont.)

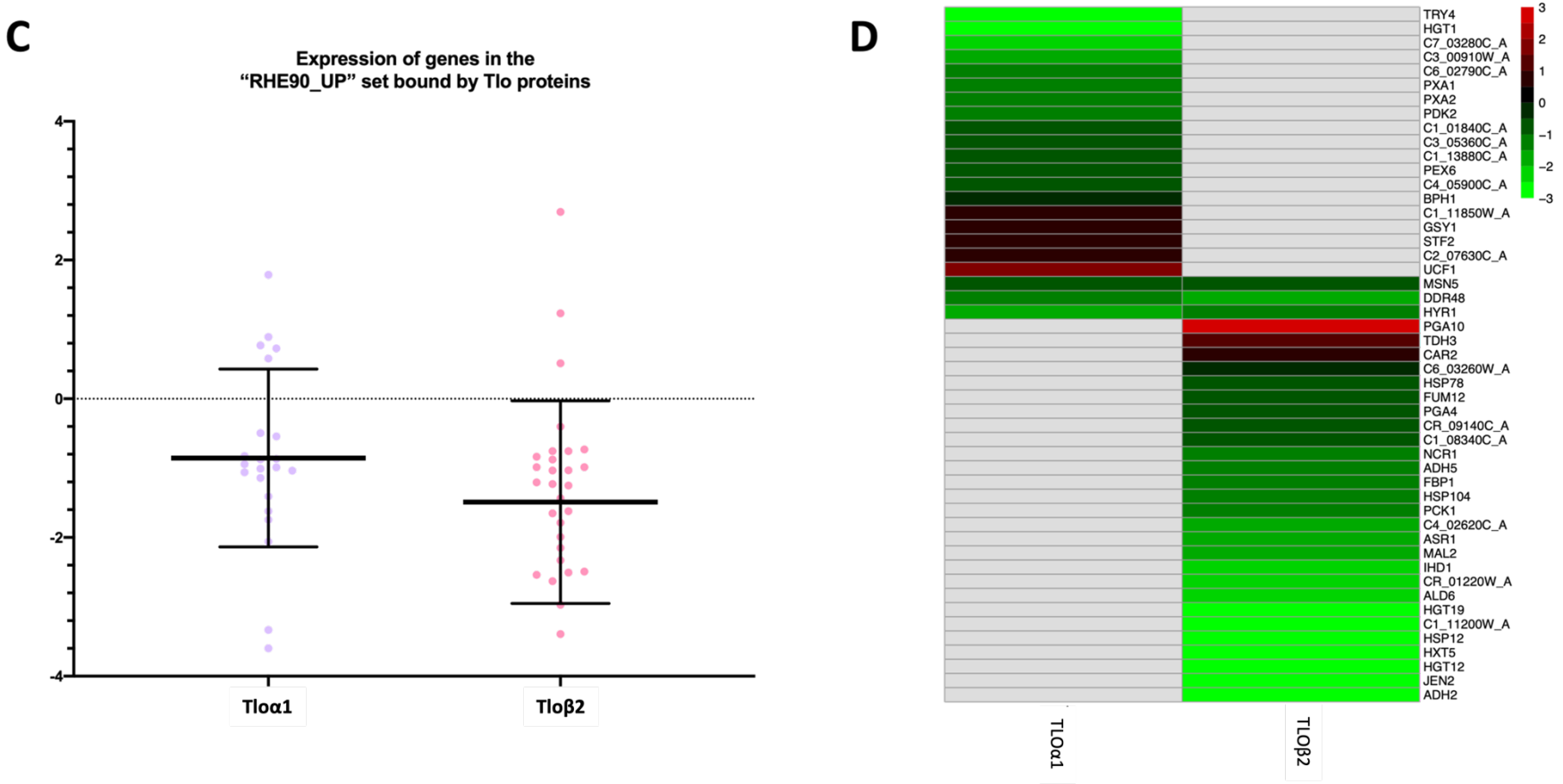


Figure 7.25 Expression of genes which Tloα1 or Tloβ2 interact with in downregulated gene sets (cont.)

Figure 7.25 Expression of genes which Tloα1 or Tloβ2 interact with in downregulated gene sets (cont.)

The L_2FC of genes that were intersected by either Tloα1, Tloβ2 peaks or both, which were significantly differentially expressed compared to the *ΔΔtlo* mutant., in groups that were highlighted by GSEA as being significantly enriched (downregulated) when intersected genes were analysed, are plotted, unpaired *t* tests with Welsh's correction were performed to determine if difference in the mean expression level were significant, (none found). Expression values for individual genes are shown in heatmaps. Heatmaps indicate the L_2FC compared to the *ΔΔtlo* mutant, with genes upregulated in red and downregulated in green (see scale on right), grey indicates that either the gene was not intersected by a peak, or was not significantly differentially expressed. (A) Expression of genes in the "KETOCONAZOLE_UP" gene set intersected by Tloα1 and Tloβ2 peaks. These are genes typically upregulated upon exposure to ketoconazole. (B) Heatmap of expression of genes in the "KETOCONAZOLE_UP" gene set intersected by Tloα1 and Tloβ2 peaks. (C) Expression of genes in the "RHE90_UP" gene set intersected by Tloα1 and Tloβ2 peaks. These are genes upregulated after exposure to reconstituted human epithelium for 90 mins. (D) Heatmap of expression of genes in the "RHE90_UP" gene set intersected by Tloα1 and Tloβ2.

intersected by peaks from both Tlos sharing the same expression patterns, Figure 7.25 (B). The *CRZ1* gene is intersected by peaks from both Tlos and is upregulated in both strains. The genes in the 'RHE90_UP" gene set which are intersected by the Tlo protein peaks are also generally downregulated, Figure 7.25 (C). With those intersected by peaks from both Tlos sharing the same regulation pattern (i.e. downregulation), Figure 7.25 (D).

7.4 Discussion

The aim of this chapter was to investigate the Tlo proteins and their interactions with DNA, either directly, or indirectly through interactions with other molecules such as RNAP or transcription factors, in the *C. albicans* cell. Two experiments were carried out, one investigating the binding of the RNA Polymerase II complex in the *ΔΔtlo* mutant and the AHY940 WT parent using an anti-RNAP antibody, and the other investigating the binding sites of the Tlo proteins in the *TLO* reintroduction strains using an anti-HA antibody that could bind the HA tag incorporated onto the C-terminal ends of the Tlos reintroduced to the *ΔΔtlo* mutant background.

7.4.1 RNA Polymerase II interactions in the ΔΔtlo *mutant and AHY940*

An experiment was performed where an anti-RNAP antibody was used to pull down the RNAP complex and investigate interaction sites on the DNA during growth in YEPD at 37 °C. The results from the RNAP ChIP analysis of the *ΔΔtlo* mutant and the AHY940 strain indicated that there were far more RNAP binding sites identified in the WT AHY940 than in the *ΔΔtlo* mutant, evident in Figure 7.2. This suggests that the *ΔΔtlo* mutant is much less transcriptionally active than the AHY940 parent. This trend mirrors what was seen in Chapter 6, where transcriptomic analysis showed that there were more genes expressed at a higher level in the WT AHY940 strain than in the *ΔΔtlo* mutant (350 genes), while fewer genes were expressed at a higher level in the *ΔΔtlo* mutant compared to AHY940 (239 genes). These data confirm the role of the Tlo/Med2 family in the enhancement of transcription rates. A summary of gene sets that RNAP interacts with in these strains which were also seen to be upregulated in transcriptomic analysis of the strains in Chapter 6 can be found in Figure 7.26.

The majority of genes intersected by RNAP peaks in the *ΔΔtlo* mutant were also intersected by RNAP peaks in AHY940, while a small number of RNAP intersected genes were unique to the *ΔΔtlo* mutant (61 genes) and the AHY940 has a much large number of unique RNAP peak intersected genes (697 genes), Figure 7.3. The majority of genes that were intersected by RNAP peaks in both strains were related to translation, one of the most basic and essential functions of the cell, and indicates that transcription may be Tlo independent. The large number of genes related to translation, specifically tRNA genes, complicated the GO analysis in these strains, and the decision to remove tRNA

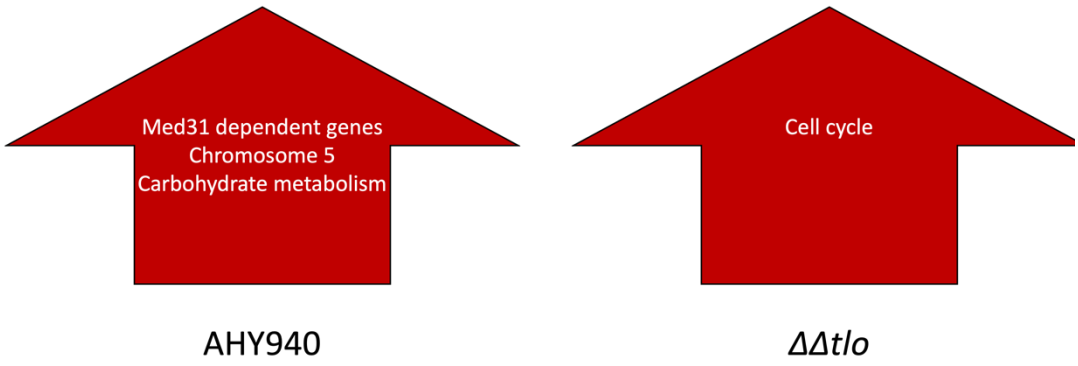


Figure 7.26 Summary of gene sets where RNAP interacts and influences expression in AHY940 and the *ΔΔtlo* mutant

A summary of gene sets in each strain over which RNAP interactions are enriched and where the expression of these gene sets is upregulated can be found above. Genes upregulated in AHY940 compared to *ΔΔtlo* mutant are on the left, and those upregulated in the *ΔΔtlo* mutant compared to AHY940 are on the right.

genes from the lists for analysis enhanced the ability to visualise other enrichments in the datasets. Genes intersected by RNAP peaks in AHY940 were enriched for carbohydrate transport, among other transport GO terms, Figure 7.4. The RNA-seq data described in Chapter 6 showed that the AHY940 strain expressed genes related to carbohydrate metabolism at a much higher level than in the *ΔΔtlo* mutant, and phenotypic data from Chapter 4 showed that the *ΔΔtlo* mutant is much less able to grow in YEPD, a glucose rich medium. These data taken together form a picture of the *ΔΔtlo* mutant strain being defective in carbohydrate metabolism, from uptake to utilisation.

The GSEA comparing the peaks in the *ΔΔtlo* mutant to those in AHY940 highlighted the trisomy of chromosome 5 in AHY940, Figure 7.5. This was also seen in the RNA-seq data in Chapter 6, and confirmed by sequencing and CHEF gels performed by collaborators, see Chapter 4. The enrichment of the "UPC2_UP" group was also found in the RNAP peak intersected genes in AHY940, but, as mentioned in Chapter 6, this may be an artefact of the chromosome 5 trisomy. Another gene set that was enriched in AHY940 was the "MED31_DN" group, which are genes that are downregulated in a *ΔΔmed31* mutant (Uwamahoro *et al.*, 2012). The enrichment of this gene set in the AHY940 WT strain, indicates that this strain is expressing genes that require Med31 (i.e. *ΔΔmed31* mutant down). The RNA-seq data in Chapter 6 also found this group enriched in the AHY940 strain. RNA-seq analysis also found that the "MED31_UP" group was enriched in the *ΔΔtlo* mutant, indicating that genes that are more highly expressed in a *ΔΔmed31* mutant (putative Med31 repressed genes) were also highly expressed in the *ΔΔtlo* mutant strain (however the *ΔΔtlo* mutant RNAP peak intersected dataset was not enriched for this gene set). The *ΔΔmed31* mutant displays similar phenotypic deficits to the *ΔΔtlo* mutant, it cannot filament, it cannot form biofilms and it is more susceptible to cell wall stress (Uwamahoro *et al.*, 2012). The "MED31_UP" gene set was also found to be enriched in the *Δtlo1/Δtlo2* mutant strain of *C. dubliniensis* described by Haran *et al.*, which also displays very similar phenotypes to the *ΔΔtlo* mutant *C. albicans* strain in here (Haran *et al.*, 2014).

Very few gene sets were enriched in the RNAP peak intersected genes in the *ΔΔtlo* mutant compared to AHY940. Two of the four sets enriched in the *ΔΔtlo* mutant strain were related to positions on chromosomes, "CA21CHRRX1", containing genes found in the first 250,000 bp of chromosome R, and "CA21CHR1X16", containing genes found between positions 2,950,000-3,250,000 bp of chromosome 1, however none of these gene sets were completely exclusively intersected by RNAP peaks in only the *ΔΔtlo*

mutant. The other two sets that were significantly enriched were the gene set containing genes in the secretory pathway biological process GO term, and the gene set identified as interacting with Erv14 in *S. cerevisiae*. The enrichment of these two gene sets would appear to be linked as both are related to the secretory pathway (Nakanishi *et al.*, 2007, Costanzo *et al.*, 2010).

In analysis of the expression level of genes intersected and those not intersected by RNAP peaks in the cells, Figure 7.6, one would expect the genes intersected by RNAP peaks to have an observably higher expression level, as these are most likely genes that are being actively transcribed. However, this was not the case. There was no statistically significant difference found in the expression of genes intersected by RNAP peaks compared to those not intersected in either strain. It does appear that the RNAP ChIP did not detect all genes that were being actively transcribed in the cells, as the RNA-seq data in Chapter 6 indicates that many more genes are being transcribed in these cells. There was no difference in the expression of genes intersected by RNAP peaks in the *ΔΔtlo* mutant and those intersected by RNAP peaks in AHY940, which was expected. In the *ΔΔtlo* mutant there was no difference in the expression of genes intersected by RNAP peaks at the promoter, at the TSS or over the ORF, however, in AHY940, the genes intersected by RNAP peaks at the promoter (including genes intersected at multiple sites) were expressed at a significantly lower level than those intersected at the TSS or over the ORF (all other intersected genes), Figure 7.7. This difference could be due to the RNAP stalling at the promoter, where the RNAP has been recruited to the promoter region, but transcription has not been activated (Wu and Snyder, 2008). RNA Polymerase II ChIP-seq has been described as an alternative to RNA-seq in fungi, where transcription of genes can be monitored in real time by measuring RNAP occupancy on genes (Tan and Wong, 2019). However, the results of the RNAP ChIP experiments discussed in this chapter indicate that further optimisation would be required for this method to replace RNA-seq.

7.4.2 Interactions of the Tlo proteins with DNA

ChIP-seq was also used to investigate the specific interactions of Tloα1, Tloβ2 and Tloγ11 with the genome. These experiments were performed in the strains generated in Chapter 5, where select *TLO* genes were reintroduced into the *ΔΔtlo* mutant background under the control of the *TET1* promoter with a 3xHA tag added to the C-terminal end of the gene. Interestingly, difficulty in detecting the Tloγ11 protein via

Western Blot in Chapter 5 was not an issue in these experiments as DNA was successfully pulled down in this strain, perhaps due to the large volume of starting material used.

The binding of the Tlos to the genome showed that while many sites and genes were intersected by peaks from all three Tlos examined, there were binding sites unique to each protein found also, Figure 7.11. Tloα1 peaks intersected the fewest number of sites in the genome, while Tloβ2 and Tloγ11 were found to have similar numbers of peaks detected. Tloα1 had around 20% of the number of peaks found for Tloβ2 and Tloγ11, and a large proportion of the genes intersected by Tloα1 peaks were also intersected by Tloβ2 and Tloγ11 peaks, possibly indicating some functional redundancy that may be a remnant of the expansion of the gene family via duplication of the ancestral locus (Dunn *et al.*, 2018). The existence of unique binding sites for each of the proteins also points toward there potentially being unique roles for each Tlo protein, which was also hinted at in the phenotypic data described in Chapter 5.

The thresholds for ChIP data reproducibility used in this work were set by the ENCODE Consortium, which aims to create an encyclopaedia of DNA elements for the human genome. This consortium works with very strict thresholds, and require that replicates share 75% of targets identified, or 80% of the top 40% of targets, if these thresholds are not met, they recommend performing another replicate (Landt *et al.*, 2012). Here, all experiments except for the Tloγ11 ChIP-seq passed this threshold, Table 7.1. One possible reason to explain this is that there may be less of the Tloγ11 protein compared to the other two, as suggested by the failure to detect this protein by Western blotting as described in Chapter 5. Another explanation may be that the Tloγ11 protein is in complex with the chromatin, discussed below, making it difficult to isolate this protein (Shiio *et al.*, 2003). However, for the purposes of this work further analysis was carried out on the samples regardless of their falling just short of the reproducibility threshold.

All three Tlo proteins were found to localise to the repetitive regions of DNA in the genome, the MRS, and the telomeres, Figure 7.10. These repetitive sequences are thought to be a hotspot for genomic recombination (Chibana and Magee, 2009). This could indicate a role for the Tlos in maintaining chromosome stability, as the binding of proteins in these regions may prevent recombination events occurring. In *C. dubliniensis,* it was found that CdTlo1 localised to the chromosome ends and repetitive sequences, and this was speculated to be involved in the maintenance of

genome stability (Haran *et al.*, 2014). The irregular chromosome structure of the *ΔΔtlo* mutant strain, Chapter 4, may be a result of the absence of Tlos in this strain, where the absence of Tlo protein binding leads to a derepression of mechanisms for chromosome stability, leading to increased levels of chromosomal rearrangement. Further investigation into the chromosome structures of the strains could help decipher if these Tlo proteins being reintroduced into the *ΔΔtlo* mutant background can stabilise the genome once again.

The majority of the Tlo binding sites in relation to genes were over the ORF, and not at the promoter region, Figure 7.11. The role of the Mediator complex in the cell is to facilitate the interaction of RNAP with DNA bound transcription factors to promote transcription, the formation of the preinitiation complex. Transcription factors generally bind upstream of the ORF in the promoter region, and this is typically where the preinitiation complex forms (Orphanides *et al.*, 1996). After the formation of the preinitiation complex, transcriptional elongation requires the RNAP molecule to move away from the preinitiation complex, termed promoter escape. Promoter escape requires the phosphorylation of the RNAP molecule, which is facilitated by Mediator. It does not appear, however, that the Mediator complex moves with the RNAP as it travels along the ORF (Soutourina, 2018). Although, ChIP-chip in *C. dubliniensis* showed that CdTlo1 localised over ORFs, while there is no free pool of CdTlo in *C. dubliniensis* (Haran *et al.*, 2014). Therefore, the interaction of Tlo proteins with ORFs could be Mediator dependent or independent. Further experiments would need to confirm this idea.

In ChIP-chip performed by Haran *et al.* identifying the interaction sites of CaTlo1 in *C. dubliniensis*, it was seen that CdTlo1 localisation could also correlate with the downregulation of genes. However, in *C. dubliniensis* there is no free pool of CdTlo1 protein, at least in the conditions that the ChIP was performed. The CdTlo1 protein only exists in complex with the Mediator (Haran *et al.*, 2014). This indicates that the Mediator complex, along with Tlos, may have a role in gene repression and not only in promoting transcription.

ChIP-seq analysis of the Tlo protein interactions with the genome highlighted that there was a clear overlap in the binding sites of the three Tlos examined, Figure 7.11. This suggests some level of redundancy in localisation (if not in function). However, there were also distinct locations intersected by peaks from each Tlo tested, and not by the

others, which lends to the idea that there are also distinct roles for each Tlo protein (or at least clade) within the cell.

The genes intersected by peaks from each of the Tlo proteins were subjected to GO analysis, Figures 7.12 & 7.13 and GSEA, Figures 7.14-7.16, to investigate the functions of the genes intersected by these protein peaks, and to determine if binding of unique gene sets was specific to any one Tlo. This analysis highlighted that genes intersected by peaks from all three Tlo proteins examined were related to translation, a basic function of the cell. There were also many genes intersected that are involved in other general biological and metabolic processes in the cell. Analysis of the genes only intersected by Tloα1 protein peaks detected enrichment for GO terms related to glycerolipid and glycerophospholipid metabolism. These pathways are important for cell signalling and transporting newly synthesised proteins to the cell membrane (Henry *et al.*, 2012) and could be a reason for the increased ability of the strains expressing *TLOα1* in the *ΔΔtlo* background to resist challenge with damaging compounds. Perhaps the response pathways are more efficient in these strains as a downstream result of Tloα1 interaction with these genes.

The genes only intersected by Tloβ2 peaks were found to be enriched for only two GO terms, nucleoside-triphosphatase activity, which is related to amino acid metabolism, and hydrolase activity. Genes that were intersected only by Tloγ11 peaks were found to be involved in transcription and general metabolism pathways. Genes that are involved in protein kinase activity were also uniquely intersected by Tloγ11 peaks. It was speculated in Chapter 5 that the strains expressing *TLOγ11* in the *ΔΔtlo* background could be deficient in protein kinase activity and that this could be an explanation for several of the phenotypes seen in these strains, such as increased sensitivity to cell wall stressors, as well as defective filamentation and biofilm formation (Navarro-Garcia *et al.*, 2005, Willger *et al.*, 2015). The unique binding of Tloγ11 at genes related to protein kinase activity could point toward disruption of the expression of these genes, and downstream consequences because of it. Tloγ11 was also shown to localise at genes involved in transcription factor activity, however the RNA-seq data in Chapter 6 showed that the transcriptome of the *ΔΔtlo::*P_{TET}*TLOγ11* strain was not very different from that of the *ΔΔtlo* mutant, indicating that, at least under the culture conditions tested, the binding of this protein at these genes does not have any great effect on the transcriptome.

The genes that each Tlo interacted with were subjected to GSEA, where the strength of the peak intersecting the gene was used as a ranking factor, Figures 7.14-7.16. Due to the difference in the number of genes intersected by Tloα1 compared to Tloβ2 and Tloγ11, far fewer gene sets were highlighted as being enriched in this dataset.

The enrichment of the "HS_UP" gene set in the list of genes intersected by Tloα1 peaks is interesting, as in the RNA-seq data in Chapter 6, the "HS_DN" gene set is enriched in the *ΔΔtlo::P_{TET}TLOα1* strain compared to in the *ΔΔtlo* mutant. The *ΔΔtlo* mutant is proposed to be in a state of chronic stress, and the enrichment of the "HS_DN" gene set in the *ΔΔtlo::P_{TET}TLOα1* strain, and the converse enrichment of the "HS_UP" gene set in *ΔΔtlo*, indicates that the expression of *TLOα1* restores a more normal, less stressed phenotype. The binding of the Tloα1 protein to the "HS_UP" gene set as found in this ChIP-seq data, indicates a potential role for Tloα1 in the repression of the transcription of these genes. In *C. dubliniensis,* it was also speculated that there is a role for Tlo proteins in both activation and repression of genes, and experiments expressing CdTlos in a *Δtlo1/Δtlo2* mutant *C. dubliniensis* showed restoration of the repression of starvation induced genes that were upregulated in the absence of Tlos (Haran *et al.*, 2014).

There also appears to be a role for Tloα1 in the activation of transcription. Genes in the "MED31_DN" and "WHITE_UP" gene sets were both shown to be sites of interaction for the Tloα1 protein, and these gene sets were both found to be enriched in the transcriptomic data comparing the *ΔΔtlo::P_{TET}TLOα1* strain to the *ΔΔtlo*. The increased expression of these genes indicates that the localisation of Tloα1 may play a role in activating their transcription. The experiments performed by Haran *et al.* in the *C. dubliniensis Δtlo1/Δtlo2* mutant also indicated a role for Tlo proteins in activating the transcription of genes, such as those related to glycolysis (Haran *et al.*, 2014). These data indicate a direct role for Tloα1 in the activation of white phase specific genes, which likely play a role in the restoration of the yeast blastospore morphology seen in the *ΔΔtlo::P_{TET}TLOα1* strain.

The enrichment of the "MED31_DN" set, coupled with its increased expression seen in the transcriptomic data, indicate that the Tloα1 protein can restore the expression of these Mediator dependent genes that are downregulated when Mediator function is compromised (i.e. by the deletion of the Med31 subunit) (Uwamahoro *et al.*, 2012). The "Med31_DN" group is also enriched in the set of genes intersected by Tloβ2 peaks

(although not statistically significant), and this gene set was found to be significantly upregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain compared to the *ΔΔtlo* mutant in RNA-seq analysis. The Tloβ2 protein was also found to bind gene groups that are impacted by disruption of Mediator function. The "MED31_UP" gene set was found to be significantly enriched in the dataset of genes that Tloβ2 interacted with, but transcriptomic sequencing indicated that this gene set is significantly downregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain compared to the *ΔΔtlo* mutant. The "MED31_UP" gene set is downregulated in the *ΔΔtlo::P_{TET}TLOα1* strain, but it was not found to be statistically significant, and the Tloα1 protein does not appear to bind these genes. It appears that Tloα1 and Tloβ2 both have a shared role in activating the expression of Mediator dependent genes, however there may be a role for Tloβ2 in repressing genes that are upregulated in the absence of Mediator that Tloα1 does not share.

The binding of the Tloα1 protein to genes that are upregulated by the Gal4 transcription factor is also noteworthy. Transcriptomic analysis found that the expression of *TLOα1* in the *ΔΔtlo* mutant background significantly increased the expression of genes bound by the Gal4 transcription factor, Chapter 6. The *GAL4* gene itself is not differentially expressed in the *ΔΔtlo::P_{TET}TLOα1* strain compared to the *ΔΔtlo* mutant, which may indicate that the interaction of Tloα1 with these genes promotes their expression independently of Gal4. Tloβ2 was also found to bind to the group of genes bound by the Gal4 transcription factor, however the transcriptomic data in Chapter 6 indicated there may be some deficiencies in the expression of the genes in this group in the *ΔΔtlo::P_{TET}TLOβ2* strain compared to the *ΔΔtlo* mutant. It was previously proposed in Chapter 6 that genes under the control of Gal4 were less expressed in the *ΔΔtlo::P_{TET}TLOβ2* strain (compared to the *ΔΔtlo::P_{TET}TLOα1* strain). The *GAL4* gene itself is not intersected by any Tloβ2 peaks, and it is not expressed highly in the *ΔΔtlo::P_{TET}TLOβ2* strain. Two genes that were highlighted as being poorly expressed in the *ΔΔtlo::P_{TET}TLOβ2* strain were *PGI1* and *TPI1*, both targets of Gal4 in the carbohydrate metabolism pathway (Askew *et al.*, 2009), and these genes were both found to be intersected by Tloβ2 peaks. It could be that Tloβ2 does not elicit the same positive effect on these genes, unlike Tloα1. While the *ΔΔtlo::P_{TET}TLOβ2* strain can grow well in YEPD with 2% glucose (as was used for phenotypic testing in previous chapters), if this strain is not expressing all genes in the carbohydrate metabolism pathways, then perhaps testing the strains at lower concentrations of glucose would highlight any deficits in the ability of the *TLOβ2* reintroduction strains to metabolise carbohydrates.

The enrichment gene sets related to cell wall morphology may indicate a role for Tloβ2 in controlling morphology of *C. albicans* cells. The "YEAST-FORM CELL WALL_BIO" group was enriched, as well as groups such as "RAS1_YEAST_UP" and "YAK1_Y_UP". While these groups are all intersected by the Tloβ2 protein peaks, transcriptomic evidence from Chapter 6 shows that these gene sets are all more highly enriched in the *ΔΔtlo* mutant strain compared to the *ΔΔtlo::*P_{TET}*TLOβ2* strain, indicating that Tloβ2 interaction may repress the expression of these genes. The "RAS1_YEAST_UP" gene set includes genes that are upregulated in yeast cells in which the adenyl cyclase regulator, *RAS1*, has been deleted. These *ΔΔras1* cells are not defective in yeast cell morphology, but form pseudohyphae in hypha inducing conditions (Harcus *et al.*, 2004). The "YAK1_Y_UP" gene set is also intersected by the Tloβ2 protein peaks, and the expression of these genes is repressed in the *ΔΔtlo::*P_{TET}*TLOβ2* strain compared to the *ΔΔtlo* mutant, again indicating a role for the Tloβ2 protein in repressing their expression. *YAK1* is a kinase required for regulation of the yeast to hyphal transition, and deletion of this gene creates a strain that is defective in this transition (Goyard *et al.*, 2008). The binding of Tloβ2 to genes that are upregulated in *ΔΔyak1* and *ΔΔras1* mutants, i.e. Yak1 and Ras1 repressed genes, and the repression of these in the *ΔΔtlo::*P_{TET}*TLOβ2* strain, indicates a role for Tloβ2 in the normal function of Yak1 and Ras1, and normal passage through the yeast-hyphal transition.

As described previously subsets of cells in the strains expressing *TLOβ2* in the *ΔΔtlo* background exhibit hyphal morphology under non-hypha inducing conditions. In Chapter 5, the RNA-seq data highlighted the increased expression of genes bound by Ume6 in the *TLOβ2* strain, which was thought to possibly contribute to this morphology. The Yak1 kinase is required for the upregulation of *UME6* during the yeast to hyphal transition (Goyard *et al.*, 2008), and as mentioned previously Tloβ2 binding appears to aid the function of Yak1. It could be the case that in the hyphal subset of cells, there is an excess of Tloβ2 binding to genes in the hyphal morphogenesis pathways which results in the altered morphology of this subset of cells.

While the transcriptomes of the *ΔΔtlo::*P_{TET}*TLOα1* and *ΔΔtlo::*P_{TET}*TLOβ2* strains were very different to that of the *ΔΔtlo* mutant, the *ΔΔtlo::*P_{TET}*TLOγ11* strain does not have many genes that were significantly differentially expressed compared to the *ΔΔtlo* mutant. This brings up the question of the role of Tloγ11 in the cell, as it does not appear to play a major role in the transcriptional response to the conditions tested. GSEA found enrichment of gene sets related to the negative regulation of biological processes in the

list of genes intersected by Tloγ11 peaks, however, comparison with the transcriptomic data from Chapter 6 did not show any change in expression of genes in these groups in the *ΔΔtlo::P_{TET}TLOγ11* strain compared to the *ΔΔtlo* mutant.

As mentioned previously, all three Tlos localised at repetitive regions of the genome, however GSEA analysis of the genes that Tloγ11 interacted with found significant enrichment at the ends of chromosomes, with the gene groups at the ends of each chromosome (apart from chromosome R) being significantly enriched, indicating that Tloγ11 localises at the telomeres and subtelomeres. Enrichment of some chromosome ends were found in the GSEA of the Tloα1 and Tloβ2 intersected genes, but not to the same extent as in the Tloγ11 intersected genes. The telomeres are repeat rich sequences, and along with the subtelomeres (~15 kbp adjacent to the telomeres) are hotspots for recombination (Dunn and Anderson, 2019). The Tlo proteins, specifically Tloγ11 were also found to interact at the MRS regions, which are also highly repetitive and prone to recombination (Dunn and Anderson, 2019). In *C. dubliniensis,* it was found that the CdTlo1 protein localised at the telomeric and subtelomeric regions of the chromosomes (Haran *et al.*, 2014). The *ΔΔtlo* mutant displayed an irregular chromosomal structure, speculated to be the result of aberrant chromosomal rearrangement, Chapter 4. It was proposed that the Tlo proteins play a role in maintaining chromosomal structure by binding at regions prone to recombination and stabilise them. While Tloβ2 was also seen to localise to these regions, Figure 7.10, perhaps this is the main role of Tloγ11 in the cell. Further experiments could be designed to investigate the role of Tlos, specifically Tloγ11 in maintaining chromosome structure in *C. albicans*, such as a serial culturing experiment that can show changes in chromosome structure over time.

7.4.3 Influence of Tlo protein binding on gene expression

It has been suggested that as the Tlo proteins are components of the Mediator complex, they have a role in controlling the expression of genes. The Tlo proteins have also been assumed to have a role independent to Mediator, and in fact the majority of Tlo protein present in the *C. albicans* cell is not associated with Mediator; so-called free Tlo (Zhang *et al.*, 2012). However, the presence of a transcriptional activation domain in the Tlo protein indicates that Tlos may potentially interact with DNA whether they are a part of the Mediator complex or not. The localisation of free Tlo to genes is proposed to augment the expression of genes by competing for binding sites with other activators,

or through the recruitment of co-activators or co-repressors (Liu *et al.*, 2016). It was seen in *C. dubliniensis* that localisation of CdTlo1 to genes was associated with both activation and repression (Haran *et al.*, 2014), and this is likely the case in *C. albicans*. Further experiments need to be performed to examine if the activation effects of Tlos are dependent on incorporation into the Mediator, perhaps by performing similar experiments to those described here in a *ΔΔmed3/ΔΔtlo* mutant strain with reintroduction of *TLOs*, as the Med3 Mediator subunit is required for Tlo integration into the complex (Zhang *et al.*, 2012).

The expression of genes that were intersected by each Tlo compared to their expression in the *ΔΔtlo* mutant was examined, Figure 7.18. It was observed that of all the genes that Tloα1 interacted with, 158 of these were found to be significantly differentially expressed in the *ΔΔtlo::*P_{TET}*TLOα1* strain compared to the *ΔΔtlo* mutant. This means that up to 29% of Tloα1 interactions could have had a direct transcriptional effect on the target genes. For Tloβ2 this figure is 17% (394 genes), and for Tloγ11 it is 0.01% (27 genes). A summary of gene sets that were bound by Tloα1 or Tloβ2 which were significantly differentially expressed can be found in Figure 7.27.

It should be noted that transcriptional activation domains are only thought to be present in the α and β clade *TLOs* and not in the γ clade genes (Liu *et al.*, 2016). This is supported by the transcriptional and phenotypic data presented in Chapters 4 and 6, where despite evidence of Tloγ11 binding at numerous genes, the expression of *TLOγ11* in the *ΔΔtlo* background does not have a dramatic effect on phenotype or transcriptome (few genes were significantly differentially expressed between the strains). The intracellular localisation of the different clades of Tlo are also interesting. While α and β clade Tlo proteins localise to the nucleus, where they can interact with the DNA, the γ clade Tlos primarily localise to the mitochondria (as well as the nucleus) (Anderson *et al.*, 2012). In this study there were no Tloγ11 binding sites identified on the mitochondrial DNA (data not shown). From these data, it is assumed that the function of Tloγ11 may not be to influence transcription, it could instead be involved in physically stabilising the genome, as discussed above, or perhaps it is involved in protein:protein interactions separate to the Mediator complex. The latter has been under investigation by Dr Dean Frawley, postdoc in the lab, who is performing mass spectrometry to identify protein interaction partners for each Tlo protein, with preliminary data indicating that the most significant interactions of Tloγ11 are not with other Mediator subunits, hinting at a role for this protein outside of Mediator-dependent transcriptional regulation.

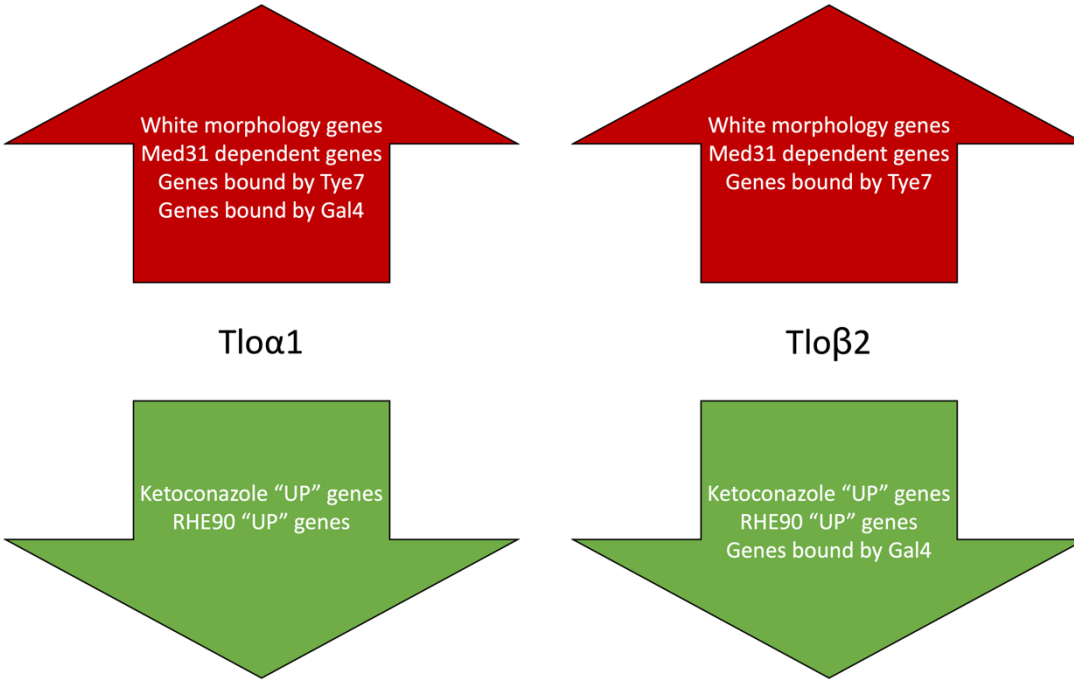


Figure 7.27 Summary of gene sets where Tloα1 or Tloβ2 interact and influence expression of genes

A summary of gene sets in each strain over which Tloα1 or Tloβ2 interactions were enriched and where the expression of these gene sets was differentially regulated compared to the *ΔΔtlo* mutant can be found above. Gene sets that were upregulated are found in the red arrows (pointing up) and those that were downregulated can be found in the green arrows (pointing down).

Investigating the effect of Tloα1 and Tloβ2 on gene expression highlighted shared and distinct genes that are influenced by either protein. Genes that were intersected by peaks from these Tlos which were also significantly up or downregulated in their respective strains compared to the *ΔΔtlo* mutant were subjected to GO analysis to determine possible functions, Figures 7.18-7.20.

Genes that were only intersected by Tloα1 peaks and upregulated in the *ΔΔtlo::*P_{TET}*TLOα1* strain were not enriched for any specific GO terms, but the genes intersected by Tloβ2 peaks and upregulated in the *ΔΔtlo::*P_{TET}*TLOβ2* strain were enriched for many GO terms, many of which were related to small molecule biosynthesis. However, the enrichment of several GO terms related to the regulation of the response to host and regulation of defence responses was interesting, specifically the binding to, and upregulation of, the *RIM101* gene. *RIM101* is a transcription factor that plays a key role in mediating the response of the *C. albicans* cell to external stimuli, mainly pH, and one of the main downstream effects of *RIM101* is stimulating the yeast to hyphal transition (Bensen *et al.*, 2004). This may be another possible explanation for the hyphal morphology seen in a subset of cells in the strains expressing *TLOβ2* in the *ΔΔtlo* background.

Upregulated genes intersected by both Tloα1 and Tloβ2 peaks were found to contain a significant number of DNA binding transcription factors, including *CRZ1, CUP9, FCR1, MIG1* and *TYE7*, Figure 7.18.

Tye7 has been described as the central transcriptional regulator for carbon metabolism, and deletion of this gene results in a strain less able to grow in glucose rich conditions in the presence of antimycin A, which inhibits respiration (Askew *et al.*, 2009). It directly activates downstream genes in the carbon metabolic pathway, genes which have also been shown to be downregulated in the *ΔΔtlo* mutant and upregulated in the *ΔΔtlo::*P_{TET}*TLOα1* and *ΔΔtlo::*P_{TET}*TLOβ2* strains. It has also been shown that a *ΔΔtye7* mutant strain is less virulent than WT. Downregulation of *TYE7* was seen in a *Δtlo1/Δtlo2* mutant strain of *C. dubliniensis* growing in YEPD, which also displayed a slower growth rate than its WT parent (Haran *et al.*, 2014). It was also seen that expression of *TYE*7 was significantly downregulated in a *ΔΔmed3* mutant *C. dubliniensis* strain. To identify if binding of Tloα1 and Tloβ2 at the *TYE7* locus requires Mediator to promote expression, candidate *TLOs* could be expressed in a *ΔΔmed3/ΔΔtlo* mutant *C. albicans* background.

In *S. cerevisiae, MIG1,* is an important regulator of glucose repression, where, in preferential high glucose conditions, genes related to alternative carbon metabolism are downregulated. This system also exists in *C. albicans,* and the *ScMIG1* homologue, *CaMIG1* is a key transcription factor involved in controlling the downregulation of alternative carbon utilisation genes (Lagree *et al.*, 2020). The upregulation of *MIG1* in the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains correlates with the improved growth rate of these strains in YEPD compared to the slower growing *ΔΔtlo* and *ΔΔtlo::P_{TET}TLOγ11* strains, which do not show high levels of expression of *MIG1*.

Regulation of high-level transcription factors involved in the glycolysis pathways and utilisation of carbohydrates indicate a direct role for Tloα1 and Tloβ2 in controlling the expression of genes in pathways essential for the utilisation of carbon sources and the overall growth and fitness of the cell.

FCR1 is a transcription factor that controls resistance to fluconazole and ketoconazole. It is a negative regulator of resistance and *ΔΔfcr1* mutants are much more resistant to these azole drugs than WT strains (Talibi and Raymond, 1999). As mentioned previously, the *ΔΔtlo* mutant strain and the *ΔΔtlo::P_{TET}TLOγ11* strain have been shown to be more tolerant of fluconazole (unpublished data), and the upregulation and expression of *FCR1* could be another contributing factor to the restoration of sensitivity of the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains (and AHY940) to azole drugs.

CRZ1 is a transcription factor in *C. albicans* that is a part of the calcineurin pathway. It is involved in cation homeostasis and virulence, as well as playing a role in azole tolerance (Karababa *et al.*, 2006). *CUP9* is a transcription factor in *C. albicans* with homology to *ScCUP9* and *ScTOS8*, which are paralogs in *S. cerevisiae* that arose from whole genome duplication. In *S. cerevisiae ScCUP9* is essential for copper tolerance and homeostasis (Knight *et al.*, 1994), and while resistance to heavy metals was not investigated in this work, it would be of interest to look at the tolerance of the *TLO* mutants to heavy metals, specifically copper, to determine if the Tlo proteins may have a role in controlling the expression of these homeostasis genes.

The genes that were intersected by Tloα1 and Tloβ2 peaks which were also downregulated in their respective strains were subjected to GO analysis to determine if any functions of gene were common. Genes that were only intersected by Tloα1 peaks

and downregulated in the *ΔΔtlo::P_{TET}TLOα1* strain compared to the *ΔΔtlo* mutant appeared to be related to carbohydrate transport, and autophagy. This trend was also seen in the genes that were only intersected by Tloβ2 peaks and downregulated in the *ΔΔtlo::P_{TET}TLOβ2* strain, and while the genes were different, they were involved in similar processes, carbohydrate transport and autophagy. The genes that were intersected by both Tloα1 and Tloβ2 peaks and downregulated in each strain were also significantly enriched for terms related to autophagy. From transcriptional analysis it appears that the *ΔΔtlo* mutant is eliciting a response like that of a starved cell. This is also the case in the *ΔΔtlo::P_{TET}TLOγ11* strain, and even though it was grown in nutrient rich medium it is unable to either detect or utilise the external nutrients, and instead induces a starvation response and catabolises itself through autophagy (Abeliovich and Klionsky, 2001). A *Δtlo1/Δtlo2* mutant *C. dubliniensis* was also described as undergoing a constitutive starvation response (Haran *et al.*, 2014). Taken together, these data indicate that the Tlo proteins may play an important role in the ability of the cells to detect and utilise nutrients, as well as repressing the transduction of aberrant signals.

Results of the gene set enrichment analysis of the genes intersected by Tloα1 and Tloβ2 peaks, and differentially expressed in the relative strain compared to the *ΔΔtlo* mutant were similar to the general trends seen throughout this analysis, while there were no enrichments found in the Tloγ11 intersected gene dataset, Figures 7.21-7.25. Tloα1 and Tloβ2 were both seen to bind, and possibly promote the expression of, white phase specific genes and genes which are downregulated upon Med31 deletion, indicating a role for these Tlo proteins in promoting the expression of these gene sets. The upregulation of genes that are typically downregulated upon deletion Med31 indicates that the reintroduction of Tloα1 or Tloβ2 plays a direct role in restoring the function of the Mediator complex. It should also be noted that many of the genes intersected by peaks from both Tlos which were found in these data sets were transcriptional regulators such as *FCR1, CUP9* and *TYE7*, which have been highlighted previously,

Both Tloα1 and Tloβ2 were found to interact with and upregulate genes bound by Tye7. The gene set bound by the Gal4 transcription factor was highlighted as being intersected by peaks from both Tloα1 and Tloβ2, however, the intersected genes only displayed increased expression in the *ΔΔtlo::P_{TET}TLOα1* strain. It was seen that many of the genes in the Gal4 bound gene set which Tloβ2 interacted with were downregulated. It could be the case that the binding of Tloβ2 to these genes is repressing their transcription, and further investigation into the relationship between the Tloα1 and Tloβ2 proteins, the

genes in the Gal4 bound gene set, and the Gal4 protein itself may highlight a difference in the effects of these two Tlo proteins. The possible implications of Tlos interacting with the Tye7 and Gal4 transcription factors has been described previously.

Gene sets that were intersected by Tloα1 and Tloβ2 peaks (and possibly downregulated by these proteins) included genes which are typically upregulated in response to ketoconazole exposure, and those which are typically upregulated after introduction to reconstituted human epithelium. These gene sets were both significantly enriched in the *ΔΔtlo* mutant strain compared to either the *ΔΔtlo::*P_{TET}*TLOα1* and *ΔΔtlo::*P_{TET}*TLOβ2* strains, indicating a role for the Tloα1 and Tloβ2 proteins in repressing the expression of these genes, perhaps independently of the Mediator complex.

The GSEA analysis indicated that Tloα1 may be involved in the expression of genes typically regulated by the Brg1 transcription factor, which was a gene set not highlighted in the GSEA analysis of the Tloβ2 peak intersected genes, perhaps hinting toward a Tloα1 specific function. Brg1 is involved in the biofilm formation pathway, and a *ΔΔbrg1* mutant is defective for normal biofilm formation (Nobile *et al.*, 2012). At 48 hours, the *ΔΔtlo::*P_{TET}*TLOα1* and *ΔΔtlo::*P_{ENO}*TLOα1* strains both formed significantly more biofilm mass than the *ΔΔtlo* mutant in an assay on plastic surface in Spider media, Figure 5.18. None of the other *TLO* reintroduction strains displayed a significantly increased level of biofilm formation compared to the AHY940 WT, apart from the *ΔΔtlo::*P_{ENO}*TLOγ11* strain, however this response is now believed to be artefactual. These data indicate a specific role for the Tloα1 protein in promoting the expression of genes related to biofilm formation. Further ChIP-seq experiments in biofilm inducting conditions could further pinpoint the role of the Tloα1 in this process.

This analysis also highlighted the specific function of Tloβ2 in interacting with and modifying the expression of genes related to the hyphal morphology. Typically, the formation of hyphae in *C. albicans* requires signal transduction of outside stimuli to promote the expression of hyphal associated genes and the repression of yeast specific genes (Basso *et al.*, 2019). What is seen in the GSEA data here is almost the opposite of this, genes related to the hyphal morphology are intersected by Tloβ2 peaks and downregulated, while those related to the yeast morphology are intersected and upregulated (although the *ΔΔtlo* mutant is pseudohyphal and could be expressing some hyphal specific genes). As mentioned previously, there is only a small subset of hyphal cells in the *ΔΔtlo::*P_{TET}*TLOβ2* strain, the rest are typical yeast blastospores. What these

data do indicate is that the Tloβ2 plays a role in the yeast-hyphal pathway, and perhaps in the hyphal cells this protein is having the opposite effect on the transcription of these genes. Further experiments on hyphal cells would help investigate the role of Tloβ2 in the hyphal morphology, and perhaps investigation into the localisation of Tloβ2 in the *ΔΔtlo::P_{ENO}TLOβ2* strain would be useful, as in this strain, a larger proportion of cells are hyphal under basal conditions, see Chapter 5.

One interesting finding from the GSEA analysis was that these two proteins also interacted with (and upregulated) genes that are typically upregulated upon oxidative stress exposure, however these experiments were performed in the absence of oxidative stress inducing agents. The oxidative stress induced gene set was not found to be enriched in these *TLO* reintroduction strains previously, and on a transcriptomic level the *ΔΔtlo* mutant strain is exhibiting a significant (uninduced) oxidative stress response compared to the *ΔΔtlo::P_{TET}TLOα1* and *ΔΔtlo::P_{TET}TLOβ2* strains. There were also differences in the number of genes in these oxidative stress gene sets that are intersected by either Tloα1 or Tloβ2 peaks, as well as the expression level of these genes in their respective strains. As seen in Chapter 5, the *ΔΔtlo::P_{TET}TLOα1* strain is more tolerant of oxidative stress induced by tBOOH than the *ΔΔtlo::P_{TET}TLOβ2* strain, Figure 5.20, however transcriptomic analysis in Chapter 6 did not unearth a transcriptional mechanism for the difference in tolerance. Perhaps a ChIP-seq experiment mapping the interactions of Tloα1 and Tloβ2 under oxidative stress conditions (and pairing with transcriptomic data) could pinpoint a role for each of these two Tlo proteins in the oxidative stress response pathway and explain the difference in tolerance in the *TLO* reintroduction strains.

7.4.5 Closing remarks

The data generated in this chapter gave valuable insights into the role of the *TLO* gene family and the Tlo proteins, in the *C. albicans* cell. Through an anti-RNAP ChIP-seq experiment, the localisation of the RNAP complex was mapped on the *ΔΔtlo* genome and the AHY940 genome and confirmed the vastly different transcriptional landscape in a cell lacking *TLO*s. Further ChIP-seq experiments pulling down Tlo proteins in the *TLO* reintroduction strains were able to map the interaction sites of these Tlo proteins to the genome. Complementation of the ChIP data with RNA-seq data from the previous chapter was able to draw some conclusions about the putative role of these Tlo proteins

in regulation of transcription. It was seen that while some binding sites and functions were shared between Tlos, there were also distinct functions identified. Tloγ11 does not appear to play a major a role in the regulation of transcription. Further investigation is required to determine if there is a role for this protein in maintaining genome stability, or perhaps it interacts with other proteins in the cell. The fact that this protein seems to exist in very low levels in the cell may also be a reason for the lack of effect seen.

The binding of Tloβ2 to genes which are involved in the yeast-hyphal switch indicates there may be a specific role for this protein in this pathway. The involvement of *TLOβ2* in the hyphal development pathway had been speculated since morphological examination of these strains found wrinkled colony morphologies on solid media and a hyphal subset of cells in liquid media under non-hyphal-inducing conditions. Tloα1 and Tloβ2 appear to share a role in positively regulating the expression of many transcription factors which are key to the health and growth of the cell. While there appears to be a specific role for Tloα1 in the formation of biofilms. Further experiments to examine the functions of other members of the α and γ clades could highlight gene specific functions, and expression of combinations of genes in the *ΔΔtlo* background will uncover more about how these proteins operate with each other in the cell.

Chapter 8

General Discussion

The expansion of the *TLO* gene family in *Candida albicans* compared to other *Candida* species (Jackson *et al.*, 2009), and compared to other eukaryotes in general, suggests a role for this gene family in its enhanced virulence, and is the basis for this study. This work utilised genetic manipulation of the genome, phenotypic testing, transcriptomic and proteomic approaches to investigate the role of the *C. albicans TLO* gene family.

In previous experiments in *C. dubliniensis,* creating a *TLO* null mutant showed that the *CdTLO* genes, of which there are two, are essential for normal cell morphology, for growth and for virulence (Haran *et al.*, 2014). In this thesis, addition to the WT *C. dubliniensis TLO* repertoire with *CaTLO*s showed that expansion of the *TLO* gene family could enhance virulence in this species, Chapter 3. These results suggest that the expansion of the *TLO* family in *C. albicans* may, at least in part, explain why this species is much more pathogenic than other *Candida* species and points toward the expansion and possible diversification of this gene family as being key to *C. albicans* being able to colonise a range of human anatomic locations and cause opportunistic infections.

One of the goals of this study was to create a *C. albicans TLO* null mutant which would allow complementation and analysis of individual *TLO* family members. A *TLO* null mutant strain of *C. albicans* would have been difficult to generate through traditional gene deletion methods, due to the size of the gene family. The presence of 14 diploid family members would have required 28 different deletion experiments, plus intermediate confirmatory steps, and marker recycling protocols between each transformation. However, CRISPR-Cas9 mutagenesis allowed for the entire *TLO* gene family to be knocked out in a single mutagenesis event, demonstrating that this method can be a useful tool for studying large gene families in *C. albicans*.

It is currently unknown if the Mediator complex remains intact in these *ΔΔtlo* mutant strains, however, data has shown that in a *C. dubliniensis Δtlo1/Δtlo2* mutant, the Mediator complex lacked the Med2, Med3, Med5, Med15 and Med16 subunits, all tail components. The same was seen in a *ΔΔmed3 C. dubliniensis* mutant (Haran *et al,* 2014). This may indicate that the Mediator complex in the *ΔΔtlo* mutant *C. albicans* strain generated here may also lack these tail subunits. However, the inclusion of a *ΔΔmed3 C. albicans* mutant in the phenotypic testing in Chapter 4 acts as a control, where phenotypic differences between the two strains are most likely due to the absence or presence of Tlos, and not defects in Mediator. The structure of the Mediator complex in the *ΔΔtlo* mutant could be analysed by tagging a separate Mediator subunit, such as

Med8, pulling down Mediator and examining the components via SDS-PAGE to determine which subunits are present (Haran *et al.,* 2014).

The *ΔΔtlo* strain of *C. albicans* generated in this work allowed for phenotypic analysis of a strain lacking all *TLO* genes and demonstrated that *TLO* genes are important for the normal function of the *C. albicans* cell. The *ΔΔtlo* mutant shows dramatic defects in morphology, Figure 4.6, growth rate, Figure 4.10, and virulence, Figure 4.16. It is less resistant to stress, Figures 4.14 & 4.15, and is unable to form true hyphae under a wide range of hypha-inducing conditions, Figures 4.7 & 4.8. RNA-seq and RNA Polymerase II ChIP-seq data suggest that the *ΔΔtlo* mutant is much less transcriptionally active than the AHY940 parent, confirming a role for the Tlo family, the Med2 subunits, for normal transcriptional function of the cell.

The *ΔΔtlo* mutant also served as a background for reintroduction of specific *TLO* family members to observe their effects in isolation, Chapter 5. Previous publications have indicated that there may be functional overlap between the *TLO* family members. This was documented in the case of the *CdTLO* genes in *C. dubliniensis*, where some traits were restored by reintroduction of either *CdTLO1* or *CdTLO2* into a *Δtlo1/Δtlo2* background. However, this research also showed that distinct traits were regulated by each gene. The two genes shared roles in restoring colony morphology and tolerance to oxidative stress, while *CdTLO1* was better at restoring hyphal growth and the growth rate in YEP-Galactose, *CdTLO2* appeared to be more involved in suppressing biofilm growth, which was increased in the *Δtlo1/Δtlo2* background compared to the *C. dubliniensis* WT (Haran *et al.*, 2014). Work in *C. albicans* by Dunn *et al.* in 2018 showed that the *C. albicans TLO*s also regulate distinct traits to some degree and proposed that the duplication of the original *TLO* locus created functional redundancy between the original and the new loci, allowing for genetic drift and new functions for the expanded family members (Dunn *et al.*, 2018).

8.1 Functions of the Tlos in *C. albicans*

Initially identified as Med2 homologues, and as such, subunits of the Mediator complex (Zhang *et al.*, 2012), the Tlo proteins play important roles in transcriptional regulation through their integration into the Mediator complex. Mediator acts to promote transcription of genes by RNA Polymerase II through assembly of a preinitiation complex at the promoter of genes, and the facilitation of subsequent phosphorylation of

the RNA Polymerase II complex that allows it to escape the promoter region and travel along the gene, resulting in transcription of the gene into mRNA (Soutourina, 2018). It has been speculated that the composition of the Mediator tail, where the Tlo proteins are located, can change the specificity of Mediator binding and direct it to different genes, and that having a large pool of Tlo proteins available that can be incorporated into the Mediator contributes to the phenotypic plasticity of this organism (Haran *et al.*, 2014).

In this work it was seen that deletion of the entire *TLO* gene family from *C. albicans* generated an unfit and less virulent strain. The transcriptome of this strain was found to be very similar to that of a *ΔΔmed31* strain, a strain in which the Mediator middle subunit Med31 had been deleted, giving rise to a cell with a dysfunctional Mediator complex (Uwamahoro *et al.*, 2012). The deletion of the Tlos seems to disrupt the function of the entire complex, indicating that the Tlo proteins as Med2 homologues are essential for the normal function of Mediator, and to the normal function of the cell. The reintroduction of Tloα1 or Tloβ2 into the *ΔΔtlo* background restored the expression of Med31 dependent genes and downregulated the Med31 repressed genes and, while there may be some Tlo specific functions, generally restoring the many of the functions of the Mediator complex. The reintroduction of Tloγ11 did not have the same effect on the restoration of Mediator function.

The binding of Mediator at genes does not however guarantee the expression of the gene. In *S. cerevisiae*, Mediator has been found to localise both at genes that are expressed and those that are not actively expressed. Gene repression by Mediator complex is due to the recruitment of the Cdk8 kinase, which can detach from the Mediator (Andrau *et al.*, 2006). The recruitment of Mediator to silenced genes has also been seen in *C. dubliniensis*. ChIP-chip experiments mapping the interactions of CdTlo1 with the genome found that this Tlo can localise to genes which are both expressed and repressed in the cell. In *C. dubliniensis*, it is proposed that CdTlo1 is only present in a Mediator associated form (in the conditions tested), and this suggests that Mediator in *C. dubliniensis* can also localise to genes that are repressed, as well as those which are actively expressed (which was also seen here, with Tloα1 and Tloβ2 both localising to genes that are up- and downregulated). It was assumed that this inactivation of Mediator in *C. dubliniensis* was also due to the recruitment of the Cdk8 kinase, similar to what was seen in *S. cerevisiae* (Haran *et al.*, 2014).

Mediator is also responsible for stimulating the phosphorylation of the RNA Polymerase II complex by the Cdk7 molecule (Kin28 in yeast), a cyclin dependent kinase associated with the preinitiation complex. The phosphorylation of the RNA Polymerase II is required for promoter escape and subsequent transcription of genes (Soutourina, 2018). There are also reports from work in human cell lines of Mediator being involved in transcriptional repression via recruitment of epigenetic regulators, specifically methyltransferases which interact with the Cdk module and which are thought to be involved in DNA methylation and histone modification, ultimately leading to gene repression (Tsutsui *et al.*, 2013). It may be that the Tlo proteins in the complex are responsible for localising Mediator to genes, however other forces influence whether or not transcription occurs.

In *C. albicans,* the majority of Tlo protein within the cell is in a 'free' form not associated with the Mediator. It is believed that there is up to 10-fold more free Tlo protein than there is in complex with the Mediator in the *C. albicans* cell (Zhang *et al.*, 2012). In contrast to this, there is no free pool of Tlo protein in *C. dubliniensis* (Haran *et al.*, 2014). This is a major difference between the *TLO* gene families in the two species. This effect does not appear to be only a gene dosage effect, as attempts to generate a free Tlo pool in *C. dubliniensis* were unsuccessful, but possibly indicates a Tlo degradation pathway present in *C. dubliniensis* to prevent build-up of free Tlo (Haran *et al.*, 2014). It has been suggested that the free Tlo in *C. albicans* may play a direct role in transcriptional regulation also, with the α and β clade proteins containing transcriptional activation domains. The localisation of Tlo protein at genes in the absence of Mediator has been proposed to exclude other activators from binding, essentially repressing the expression of these genes (Liu *et al.*, 2016). Some phenotypes were only seen in strains expressing the *TLO* genes under the strong *ENO1* promoter, such as increased virulence in the *ΔΔtlo::*P_{ENO}*TLOα1* strain and increased hyphal formation in the *ΔΔtlo::*P_{ENO}*TLOβ2* strain, discussed below, indicating that these phenotypes may rely on generation of a free pool of Tlo protein.

Further ChIP-seq experiments could be designed to examine the localisation of the Mediator complex in relation to Tlo protein in these strains to determine if Mediator is localising to all expressed genes, which would indicate if Mediator incorporation is required for gene activation by Tlo proteins. This could be done by tagging another Mediator subunit, such as one in the middle or head domains. These experiments would

also reveal if Tlo binding in complex with Mediator localises to repressed genes in *C. albicans*, and if free Tlo protein can repress gene transcription independently.

The Tlo proteins may also play a role in maintaining genome stability, binding to the DNA at repeat rich regions where recombination is prone to occur and controlling aberrant recombination events. The telomeres, subtelomeres, long terminal repeats and the MRS regions of the *C. albicans* genome are all repeat rich and have been shown to be recombination hotspots (Dunn and Anderson, 2019). In *C. dubliniensis* it was shown that the CdTlo1 protein localised to the chromosome ends and was proposed to play a role in maintaining the integrity of the genome (Haran *et al.*, 2014). Here, in *C. albicans,* the Tlo proteins, particularly Tloγ11, and to a lesser extent Tloβ2, in *C. albicans* were found to localise to these regions, including the telomeres, centromeres and MRS regions, see Figure 7.10. It is proposed that this localisation may play a role in maintaining the structure of the genome. Some observations have suggested that the deletion of the *TLO* gene family resulted in the *ΔΔtlo* mutant displaying an irregular chromosome structure thought to be a result of chromosomal shuffling. Further investigation into this potential role for the Tlo proteins could be investigated through serial culturing of strains to examine the karyotype stability over time, and to determine if the absence of Tlos leads to increased chromosomal rearrangement, and if the reintroduction of single, or multiple Tlos can restore stability.

There is also the possibility that Tlo proteins interact with other proteins in the cell causing post-transcriptional effects. Mass spectrometry experiments currently being conducted in the lab are attempting to identify protein:protein interaction partners for the Tlos in WT *C. albicans* cells. These experiments have found that the Tlo proteins have significant interactions with proteins involved in ribosome synthesis and maintenance, amino acid synthesis, protein translation and protein degradation, as well as mitochondrial maintenance and function, ER maintenance, transport of macromolecules and amino acids, general gene transcription and replication (Dean Frawley, personal communication). These findings may indicate a role for the Tlo proteins in protein synthesis, modification and transport. Studies in *S. cerevisiae* have shown that the ScMed2 protein interacts with proteins related to the movement of proteins into and between the Golgi and the endoplasmic reticulum, Get1 and Get2 (Pan *et al.*, 2006, Collins *et al.*, 2007). The Tlo proteins were also found to interact with proteins involved in the response to external stresses, ergosterol biosynthesis, filamentation and chromatin remodelling. Studies in *S. cerevisiae* have also shown

interaction of the ScMed2 protein with proteins required for oxidative stress response, Yap1, and general stress responses, Rpn4 and Whi2, with the Erg6 protein involved in ergosterol biosynthesis, and with proteins that are important in chromatin remodelling, Htl1, Itc1, Snf5 and Swi3 (Collins *et al.*, 2007). All of these are pathways that were affected by the deletion of the *TLO* genes from *C. albicans* and which were restored by reintroduction of Tloα1 or Tloβ2. It could be that the Tlo proteins act through modifying transcription through interaction with DNA and also through protein:protein interactions to enact their functions within the cell.

8.2 Tlo proteins have common and distinct functions

The experiments described in Chapters 3, 5, 6 and 7 were aimed at uncovering specific roles for individual *TLO* genes in the *C. albicans* cell and determining if individual *TLO*s had redundant or distinct functions that would lend to the idea that the expansion and diversification of the *TLO* family plays a part in the plasticity of the *C. albicans* response to external stimuli. For the purposes of this discussion, the results of expressing the *TLOγ11* gene in the *ΔΔtlo* background will be discussed separately to *TLOα1* and *TLOβ2* due to the lack of clear effects of *TLOγ11* expression on phenotype and transcriptome.

Knocking out the entire *TLO* gene family in *C. albicans* highlighted the importance of this gene family in the normal function of the cell. This mutant strain was much less fit and much less virulent than its parent, WT *C. albicans* AHY940. The *ΔΔtlo* mutant *C. albicans* phenocopies the *Δtlo1/Δtlo2 C. dubliniensis* mutant described by Haran *et al.* 2014, indicating that the expansion of the *TLO* family in *C. albicans* is a key differentiator between the two species (Haran *et al.*, 2014). As stated previously, the *C. albicans ΔΔtlo* mutant exhibited a transcript profile that resembled that of a *ΔΔmed31* mutant with a defective Mediator (Uwamahoro *et al.*, 2012).

The Tloα1 and Tloβ2 proteins were revealed to regulate many of the same genes, enacting common functions. While the *ΔΔtlo* mutant is an unfit, pseudohyphal and less virulent strain, reintroduction of the *TLOα1* or the *TLOβ2* gene into this background, under either a weak or a strong promoter, could restore many of the mutant's phenotypes. Phenotypic testing found that expression of either these two genes in the *ΔΔtlo* background could restore growth rates, Figure 5.14, biofilm formation, Figure 5.18, resistance to cell wall stressors, Figure 5.19, oxidative stress, Figure 5.20 and virulence, Figure 5.21. These genes could also restore the yeast blastospore

morphology, however the expression of *TLOβ2* in the *ΔΔtlo* background also generated a subset of cells with a hyphal morphology in non-hypha-inducing conditions, which appeared to be gene dosage dependent, with expression of *TLOβ2* under the strong *ENO1* promoter creating a larger subset of hyphal cells, Figures 5.5 & 5.9. RNA-seq analysis identified that there were transcriptional pathways affected by the expression of these *TLO*s that are probably responsible for many of the phenotypes observed, such as those related to carbohydrate metabolism, those upregulated in white phase cells, and genes up or downregulated during hyphal growth among others. When the expression data sets were coupled with data from the ChIP-seq experiments to map the interactions of the Tlo proteins with the genome, it was seen that the Tloα1 and Tloβ2 proteins interacted with the genes involved in these pathways and were likely directly impacting control of their expression.

8.2.1 White and opaque morphologies

The *ΔΔtlo* mutant was found to express many genes related to the opaque cell type, while the WT strain expressed genes typically upregulated in white cells, indicating that the presence of the *TLO* family possibly plays a role in the control of the switch between white phase and opaque phase cell types. The Tloα1 and Tloβ2 proteins were found to localise to genes which are typically upregulated in white cells, and these genes were also found to be upregulated in these strains compared to the *ΔΔtlo* mutant. The white phase *C. albicans* cell is the typical morphology of this species, the mating competent opaque phase cell is rarely observed *in vivo*. While switching is stochastic, external stimuli can greatly increase the rate of white to opaque switching, such as oxidative stress, exposure to DNA damaging agents and exposure to hydroxyurea (Takagi *et al.*, 2019, Lohse and Johnson, 2009). The AHY940 WT strain with all *TLO*s intact was also observed to express white phase specific genes at a much higher level than the *ΔΔtlo* mutant, indicating that the *ΔΔtlo* mutant is not exhibiting typical white phase morphology, and in fact appears to be, transcriptionally, more opaque-like, although this is possibly masked by the pseudohyphal morphology. Future experiments could be designed to investigate whether the *ΔΔtlo* mutant is in fact opaque, perhaps using electron microscopy to compare the mutant cells to typical opaque cells. However, from the data generated here, it appears that expression of white phase specific genes, mediated by the Tlo proteins, appears to be key in maintaining typical white phase yeast morphology.

8.2.2 Carbohydrate metabolism

From phenotypic analysis, it was seen that the *ΔΔtlo* mutant grew significantly more slowly in YEPD at 37 °C than the WT AHY940 strain, Figure 4.10, and transcriptomic analysis indicated that the *ΔΔtlo* mutant was expressing genes related to carbohydrate metabolism at a much lower level than the WT, Figure 6.6. Expression of *TLOα1* and *TLOβ2* in the *ΔΔtlo* background restored the expression of genes involved in carbohydrate metabolism, Figure 6.15, and restored growth rate in YEPD, Figure 5.14. Tye7 is an important transcriptional regulator that controls the expression of genes in the carbohydrate metabolism pathway (Askew *et al.*, 2009). The Tloα1 and Tloβ2 proteins localised to the *TYE7* gene and appeared to promote its expression. These Tlo proteins also interacted with genes regulated by Tye7. It is proposed that the interaction of Tloα1 and Tloβ2 at these sites promotes the expression of these genes and allows for optimised growth in carbohydrate-rich media. Investigation into the relationship between Tye7 and the Tlos on a protein level would help to determine if interaction between these proteins also aids the activation of genetic expression. The gene set bound by the Gal4 transcription factor was also highlighted in this analysis, and it was observed that both the Tloα1 and Tloβ2 proteins localised to genes regulated by Gal4, Figure 7.24. The genes bound by Tloα1 were generally upregulated, however many of the genes that Tloβ2 interacted with were downregulated. While Gal4 is a transcription factor involved in controlling the expression of genes in carbon metabolism pathways (Askew *et al.*, 2009), there was no difference observed in the ability of the two strains to grow in YEPD or in YEP-Gal. This could indicate that while the Tloα1 and Tloβ2 proteins are both involved in this pathway, the net effect of the proteins in the conditions tested were the same. Potential differences in the impact of each Tlo could be investigated via growth in media with lower levels of glucose, or media containing alternative carbon sources. Expression of both these proteins in the *ΔΔtlo* background, with unique protein tags for discriminatory ChIP-seq, could also help uncover if there is competition between the two proteins as activators and repressors in this pathway. This data indicates that both of these Tlos, and likely more, are required for full functionality of the cell, where each Tlo has specific roles in controlling these pathways and that one Tlo can pick up the slack, but the expanded repertoire is required for optimal function.

8.2.3 Sterol biosynthesis and azole tolerance

There is also a proposed role for the *TLO* family in maintaining cell wall sterol content and azole sensitivity. The *ΔΔtlo* mutant has been shown to display an increased tolerance to azole drugs, believed to be due to the altered sterol content in the cell membrane. The cell membrane of the mutant lacks typical sterols, specifically ergosterol, and instead contains many toxic intermediates, suggesting that the normal ergosterol biosynthesis pathway is affected (Dr James O'Connor-Moneley, personal communication). The lack of ergosterol in the membrane creates a cell which is more tolerant to azole drugs, as the target of azole drugs, Erg11, is not present. The *TLO*s appear to play a role in normal ergosterol biosynthesis, both through transcriptional control of the ergosterol biosynthesis pathway, and through a direct role of the Tlo proteins in interacting with proteins involved in these pathways, as uncovered by mass spectrometry analysis (Dr Dean Frawley, personal communication). The *ΔΔtlo* mutant expresses genes which are typically upregulated during exposure to azole drugs. The transcriptome of this strain also matches that of an *ΔΔupc2* mutant, with the *UPC2* gene itself also being repressed. Upc2 is an important transcriptional factor controlling the expression of genes in the ergosterol biosynthesis pathway (Vasicek *et al.*, 2014). Under drug-free growth conditions, Upc2 deletion does not appear to influence the expression of *ERG* genes. In a *ΔΔupc2* mutant these genes are still expressed under these conditions. However, azole exposure typically increases the expression of *ERG* genes, including *ERG11*, via Upc2 binding to promoters and activating transcription. Overexpression of *ERG11* is a known mechanism enlisted by *C. albicans* to overcome challenge with azole drugs. This is abolished in a *ΔΔupc2* mutant, with this strain displaying increased sensitivity to azole drugs (MacPherson *et al.*, 2005). The lack of *ERG11* expression in drug-free conditions in the *ΔΔtlo* mutant, and the lack of ergosterol in the cell membrane, combined with the downregulation of *UPC2* could be responsible for the increased tolerance to azoles seen in this strain. If *UPC2* was expressed in the cell, would exposure to fluconazole increase the expression of *ERG11*? Or is Tlo protein essential to the function of the Upc2 transcription factor? This could be investigated by artificially increasing the expression of *UPC2* in the *ΔΔtlo* mutant and testing azole tolerance. RNA-seq and ChIP-seq could be performed in the presence of fluconazole to further investigate the role of the Tlos in the azole drug response.

The reintroduction of either *TLOα1* or *TLOβ2* into the *ΔΔtlo* background can rescue ergosterol biosynthesis, resulting in restored sensitivity to azoles. Though the Tlo proteins do not localise to the *ERG11* gene directly, its expression appears to be activated in some way by the reintroduction of *TLOα1* or *TLOβ2* into the *ΔΔtlo*

background. This suggests that the *TLO*s are essential for synthesis of cell membrane sterols through the ergosterol biosynthesis pathway. The Tloα1 protein was found to bind to genes repressed by Upc2, and these genes were downregulated in this strain. The Tloα1 protein does not localise to the *UPC2* gene itself, but the expression of *UPC2* is significantly higher than in the *ΔΔtlo* mutant. Similar results were also observed in the Tloβ2-expressing strain, however, the enrichment of the gene set repressed by Upc2 fell just short of the significance threshold. The Tloβ2 protein does localise to the *UPC2* gene, and while the expression of the *UPC2* gene is increased compared to the *ΔΔtlo* mutant, this difference was not statistically significant. There could also be a role for the Tlos interacting on a protein:protein level with Upc2 to aid its function, which could be investigated by protein pulldown experiments and mass spectrometry. These data indicate a role for Tloα1 and Tloβ2 in the expression of *UPC2* and *ERG11*, and indicate that they play an essential part in cell membrane synthesis and integrity, as well as in mediating the response of cells to azole drug exposure.

Other genes involved in drug tolerance in *C. albicans* were differentially expressed in the reintroduction strain compared to the *ΔΔtlo* mutant. *TAC1*, a transcription factor which promotes the expression of drug resistance genes, was expressed significantly higher in the *ΔΔtlo* mutant than in either of the *ΔΔtlo::*P_{TET}*TLOα1* or *ΔΔtlo::*P_{TET}*TLOβ2* strains (data not shown) (Coste *et al.*, 2004). However, one of the main downstream targets of Tac1, *CDR1,* was expressed more highly in the reintroduction strains than in the *ΔΔtlo* mutant (data not shown). Typically, higher levels of expression of *CDR1* correlate with increased levels of resistance to fluconazole (Coste *et al.*, 2004). This could indicate a role for the Tlo proteins in controlling the expression of, and the action of, the *TAC1* transcription factor, which could also impact the ability of the cell to tolerate azole drugs.

8.2.4 Stress responses

The *ΔΔtlo* mutant strain appears to be in a state of chronic stress, indicated by the significant upregulation of stress response genes in this strain compared to the WT. Genes upregulated include those typically induced upon exposure to oxidative stress, osmotic stress and upon encountering host immune cells. Genes related to autophagy are also upregulated, believed to be a starvation response to the inability to adequately utilise external nutrients. The reintroduction of *TLOα1* or *TLOβ2* results in the restoration of the normal expression of many of these gene sets, with the Tlo proteins

found localising to these genes in ChIP-seq experiments. This indicates the Tlo proteins likely play a role in transcriptional repression and appear to play a role in maintaining cellular homeostasis. The proteins were also found to localise to and repress genes that are typically upregulated in response to exposure to reconstituted human epithelium. This gene set was constitutively upregulated in the *ΔΔtlo* mutant and indicates a role for the *TLOs* in controlling the response to host cells. Further experiments would enable investigation into the role of these *TLO*s in the activation of host defence responses, which was highlighted as being a significantly enriched group of genes in the genes that were bound by Tloβ2, perhaps indicating differences in the abilities of these strains to respond to host defences.

This analysis also highlighted the fact that the Tloα1 and Tloβ2 proteins both localised to a subset of oxidative stress response genes and upregulated them, although these interactions did not appear to induce detectable differential expression of these genes in the respective strains on a transcriptome wide level when compared to the *ΔΔtlo* mutant. The *TLOα1* and *TLOβ2* reintroduction strains did exhibit different levels of tolerance to oxidative stress, Figure 5.20, which was not fully explained by the transcriptomic analysis described in Chapter 6. However, if a ChIP-seq experiment was performed under oxidative stress conditions, the interactions of the Tlo proteins with these genes could be investigated, and perhaps differences in the actions of the proteins could be investigated.

8.2.5 Tlo specific functions

There were some specific functions of Tlos highlighted in this study. In the case of Tloα1, data suggest that this protein is involved in the biofilm formation pathway through its interaction with the set of genes typically bound by the Brg1 transcription factor (Nobile *et al.*, 2012). From the data generated, it appears that the interaction of Tloα1 with these genes promotes their expression. Phenotypic data suggest that the strains expressing *TLOα1* in the *ΔΔtlo* background were able to form more biofilm than the *ΔΔtlo* mutant, the *TLOβ2* and *TLOγ11* reintroduction strains (excluding *ΔΔtlo::*P_{ENO}*TLOγ11* strain) and even the AHY940 WT after 48 hours in Spider medium on plastic plates, Figure 5.18. The biofilm formation pathway is relatively recently evolved (Nobile *et al.*, 2012), and it could be that the function of *TLOα1* in the pathway evolved with it over time as the *TLO* family expanded and functions diverged among family members. As above, it would be interesting to investigate this mechanism further, to perform RNA-seq and ChIP-seq

experiments on the strains growing in biofilm inducing conditions to determine the role of the Tloα1 protein, and to determine if Tloα1 interacts with Brg1 at a protein level to promote the expression of these genes.

There also appears to be a role for *TLOα1* in promoting the virulence of strains in the *G. mellonella* model, where the *ΔΔtlo::*P_{ENO}*TLOα1* strain was seen to be the most virulent, see Figure 5.21. It has previously been described that *TLOα1* was highlighted as a gene that is specifically upregulated during infection of *G. mellonella*, but one that is not upregulated during murine infection (Amorim-Vaz *et al.*, 2015). The fact that this result is only seen in the strain expressing the gene under the strong promoter could indicate that this phenotype is a result of generation of a free pool of Tloα1. Further experiments, including RNA-seq and ChIP-seq under these conditions, could identify the specific role that *TLOα1* plays in this system, and further work could indicate whether other *TLO*s play a role in virulence in a murine infection model if *TLOα1* does not.

The *TLOβ2* gene also appears to have a specific role in controlling hyphal morphogenesis in *C. albicans*. Both the *ΔΔtlo* mutant *C. albicans* strain and a *Δtlo1/Δtlo2* strain of *C. dubliniensis* were unable to form hyphae under a wide range of hypha-inducing conditions(Haran *et al.*, 2014). However, a *ΔΔmed31* mutant strain of *C. albicans* can still form hyphae under specific conditions, albeit at a delayed rate (Uwamahoro *et al.*, 2012). In this strain the *TLO* genes are still present, which may also indicate the importance of these genes in hyphal morphogenesis. Reintroduction of *TLOα1* into the *ΔΔtlo* background restored WT blastospore morphology, Figure 5.6, and allowed cells to form hyphae under inducing conditions, Figure 5.7, but when the *TLOβ2* gene was expressed in the *ΔΔtlo* background, a hyphal subpopulation of cells was seen in non-hypha-inducing conditions in liquid cultures, and the colonies displayed wrinkled morphologies on solid non-hypha-inducing media, Figure 5.6. This hyphal morphology in non-hypha-inducing conditions was seen in a much greater percentage of cells in the *ΔΔtlo::*P_{ENO}*TLOβ2* strain, Figure 5.10. When the *CaTLOβ2* was expressed in the WT *C. dubliniensis* background, similar morphologies were seen, again with a larger amount of hyphal growth seen in the strain expressing the gene under a stronger promoter, Figure 3.3. The Tloβ2 protein was also found to localise to, and modulate the expression of, many genes in the hyphal morphogenesis pathway. While the expression patterns of the genes bound was correlated to the yeast morphology, it could be speculated that the increased binding of these genes and the ability of this protein to modulate their expression could explain the hyphal morphology of the subpopulation

of cells. This could prime the cells for the shift to hyphal growth, which was seen to be expedited upon induction with serum and temperature shift in the *ΔΔtlo::P_{TET}TLOβ2* strain and to more of an extent in the *ΔΔtlo::P_{ENO}TLOβ2* strain, Figure 5.10, which could be a result of the increased level of protein present in the latter strain, or the generation of a pool, or indeed a larger pool, of free Tlo protein. As suggested with the role of Tloα1 in the biofilm formation pathway, the function of *TLOβ2* in the hyphal morphogenesis pathway may have developed alongside the morphological pathway, where this gene developed a significant relationship with the pathway as it evolved to be more efficient in *C. albicans* (Moran *et al.*, 2012). While other Tlos, such as *TLOα1*, can still function to trigger the morphological change, likely a function conserved during expansion, and that *C. dubliniensis* with far fewer *TLO*s can still form hyphae, less efficiently under specific, stricter, inducing conditions (Moran *et al.*, 2012), *TLOβ2* has evolved a specific role in this pathway over time. Further ChIP-seq and RNA-seq experiments in these strains could reveal the role of *TLOβ2* in this morphological pathway, perhaps under hyphal inducting conditions, or by examination of the hyphal subpopulation under non-inducing conditions.

Expression of the *TLOγ11* gene appeared to have had little to no effect on the phenotype of the *ΔΔtlo* mutant, nor on the transcriptome, and while the ChIP-seq data shows enriched binding to gene sets related to the negative regulation of transcription, these genes were not differentially transcribed in the *ΔΔtlo::P_{TET}TLOγ11* compared to the *ΔΔtlo* mutant. Previous questions have been raised about the presence of Tloγ11 protein in the cell since this protein was undetectable by Western Blot, regardless of the level of mRNA expressed. A *TLOγ5* gene was also introduced to the *ΔΔtlo* mutant background under a strong *ENO1* promoter and exhibited a similar lack of phenotypic effect (data not shown). A mutant strain in an SC5314 WT background, which had all other *TLO* genes deleted apart from *TLOγ5,* also displayed the same phenotypic characteristics as these strains. This could indicate a trend in the γ clade, although only two of the seven γ clade genes have been tested. The unusual phenotype exhibited by the *ΔΔtlo::P_{ENO}TLOγ11* strain in biofilm assays may in fact be an artefact of some kind, and the fact that an N-terminal HA tagged *TLOγ11* gene expressed under the *ENO1* promoter in the *ΔΔtlo* mutant background did not confer this phenotype also strengthens this possibility. One trend that remained constant with the γ clade was that none of the tagged proteins could be visualised by Western Blot, pointing toward a clade wide trait. It could be that the proteins are not being translated, that the alternative splicing of γ clade transcripts affects the translation of the tag, that the proteins are being rapidly

degraded after translation, or that the proteins are in complex with the chromatin, making isolation difficult, described below.

It was not possible to generate a free pool of CdTlo1 in *C. dubliniensis* by artificial overexpression of the gene, however it was possible to generate a free pool of CdTlo2, as well as a free pool of CaTloα12 and of CaTloβ2 in *C. dubliniensis* through heterologous expression (Liu *et al.*, 2016). In a *C. dubliniensis ΔΔmed3* background there was a large decrease in CdTlo1 protein, with no decrease in mRNA levels detected. By treating that same strain with a proteosome inhibitor the authors could produce a pool of CdTlo1 protein. This led the authors to conclude that there was an intrinsic property to the sequence of CdTlo1 that predisposed it for degradation by the proteasome via ubiquitylation if it was not protected by integration into Mediator. It is possible that the Tloγ11 protein in *C. albicans* may also be unstable because it appears to be a target for degradation. The upregulation of genes related to the proteasome complex in the *ΔΔtlo::P_{TET}TLOγ11* strain which was not seen in any other *TLO* reintroduction strain also indicates that there may be a specific property of the Tloγ11 protein that is promoting its own degradation. It could be that the data gathered on the *ΔΔtlo::P_{TET}TLOγ11* strain in RNA-seq experiments did not reveal much transcriptional difference in this strain than in the *ΔΔtlo* mutant due to the physical lack of protein present in the cell to enact transcriptional change. Although Tloγ11 protein was successfully pulled down via an anti-HA antibody for the ChIP-seq experiments, this is a highly sensitive method with a large input of starting material. The interactions detected for this protein are likely representative of all interactions of the protein in all the cells in the culture, and not of the interactions taking place consistently in each individual cell. This could be the reason for the decreased levels of reproducibility between *ΔΔtlo::P_{TET}TLOγ11* RNA-seq and ChIP-seq samples seen. A possible experiment to investigate this further would be to treat the cells with a proteosome inhibitor prior to protein extraction and to attempt to detect the Tloγ11 protein via Western Blot.

Although it was not detectable using Western blotting, Tloγ11 protein was successfully pulled down in the ChIP-seq experiments and in mass spectrometry experiments performed by Dr Dean Frawley (personal communication), indicating some level of protein present in the cells. Mapping of the protein interactions across the genome did reveal increased Tloγ11 binding at the chromosome ends and at MRS regions of the genome. As discussed above, it is thought that the binding of Tlo proteins to these regions may play a role in maintaining chromosomal stability. Perhaps the intimate

relationship between Tloγ11 and chromatin prevented successful isolation during routine protein extraction. Chromatin associated proteins are typically expressed at low levels and are difficult to extract (Shiio *et al.*, 2003), and that these proteins may only be detectable when high sensitivity methods, such as immunoprecipitation, are used.

The specific roles for each Tlo also bring into question the interaction between the Tlos in the WT cell. It is known that the *TLO* genes are under various levels of transcriptional control, such as telomere positional effect, and their expression is noisy (Anderson *et al.*, 2014), which likely means that each Tlo is present at different levels in individual cells. The balance between Tlos and their functions could have different phenotypic, transcriptional and morphological outcomes. For instance, if *TLOα1* and *TLOβ2* were both expressed in the *ΔΔtlo* background would the hyphal subpopulation still exist, and would the colonies be smooth or wrinkled on solid media? Could the presence of Tloγ11, or any other Tlo, prevent the interactions of other Tlos with their genomic targets and prevent the promotion of gene transcription? These questions could be answered with further experimentation, perhaps including the expression of many combinations of *TLOs* in the *ΔΔtlo* background. There is also the question of the distinct states of Tlo proteins, either Mediator bound, or free Tlo, which should also be investigated further, perhaps through deletion of the Med3 subunit, which is required for Tlo integration into Mediator (Zhang *et al.*, 2012), and the use of a *ΔΔtlo/ΔΔmed3* mutant background to reintroduce *TLOs* and determine if the effects seen in this work are reliant on Tlo incorporation into the Mediator.

8.3 Conclusions and future directions

This work has confirmed the hypothesis that the *TLO* gene family is important to the normal function and virulence of *C. albicans*, and that expansion of the *TLO* repertoire can enhance fitness and virulence. While further experiments need to be carried out, such as investigation of the specific roles of all other *TLO* genes, examining the requirement of Mediator incorporation for Tlo function and determining the relationships between the Tlos themselves, this work has shown that *TLO*s have common and unique functions which indicates that having such an expanded repertoire is likely to aid the plasticity of the *C. albicans* phenotype which is key to its increased ability to colonise and cause infection in humans.

Having a more versatile Med2 subunit pool available may allow the cell to adapt quickly to changing environmental stimulus through incorporation of these subunits into the Mediator complex, as well as the free Tlo effects that these proteins could have in isolation. The expansion of the *TLO* gene family in *C. albicans* appears to directly contribute to the increased virulence of this species compared to other *Candida* species and may explain, at least in part, why this yeast species is more successful than its close relatives in occupying its niche as a human commensal and opportunistic pathogen.

Bibliography

ABACI, O., HALKI-UZTAN, A. & ATES, M. 2008. Specific identification of *Candida albicans* and *Candida dubliniensis* by PCR using species-specific primers. *Annals of Microbiology,* 58(2), 325-331.

ABELIOVICH, H. & KLIONSKY, D. J. 2001. Autophagy in yeast: Mechanistic insights and physiological function. *Microbiology and Molecular Biology Reviews,* 65(3), 463-479.

AL-RUSAN, R. M., DARWAZEH, A. M. G. & LATAIFEH, I. M. 2017. The relationship of *Candida* colonization of the oral and vaginal mucosae of mothers and oral mucosae of their newborns at birth. *Oral Surgery Oral Medicine Oral Pathology Oral Radiology,* 123(4), 459-463.

ALBY, K. & BENNETT, R. J. 2009. Stress-Induced Phenotypic Switching in *Candida albicans*. *Molecular Biology of the Cell,* 20(14), 3178-3191.

AMORIM-VAZ, S., TRAN, V. D. T., PRADERVAND, S., PAGNI, M., COSTE, A. T. & SANGLARD, D. 2015. RNA Enrichment Method for Quantitative Transcriptional Analysis of Pathogens *In Vivo* Applied to the Fungus *Candida albicans*. *Mbio,* 6(5), e00942-15.

ANDERSON, M. Z., BALLER, J. A., DULMAGE, K., WIGEN, L. & BERMAN, J. 2012. The Three Clades of the Telomere-Associated *TLO* Gene Family of *Candida albicans* Have Different Splicing, Localization, and Expression Features. *Eukaryotic Cell,* 11(10), 1268-1275.

ANDERSON, M. Z., GERSTEIN, A. C., WIGEN, L., BALLER, J. A. & BERMAN, J. 2014. Silencing is noisy: population and cell level noise in telomere-adjacent genes is dependent on telomere position and Sir2. *PLoS Genet,* 10(7), e1004436.

ANDERSON, M. Z., WIGEN, L. J., BURRACK, L. S. & BERMAN, J. 2015. Real-Time Evolution of a Subtelomeric Gene Family in *Candida albicans*. *Genetics,* 200(3), 907-919.

ANDRAU, J. C., VAN DE PASCH, L., LIJNZAAD, P., BIJMA, T., KOERKAMP, M. G., VAN DE PEPPEL, J., WERNER, M. & HOLSTEGE, F. C. P. 2006. Genome-wide location of the coactivator mediator: Binding without activation and transient Cdk8 interaction on DNA. *Molecular Cell,* 22(2), 179-192.

ANSARI, S. A., GANAPATHI, M., BENSCHOP, J. J., HOLSTEGE, F. C. P., WADE, J. T. & MORSE, R. H. 2012. Distinct role of Mediator tail module in regulation of SAGA-dependent, TATA-containing genes in yeast. *EMBO Journal,* 31(1), 44-57.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M., SHERLOCK, G. & GENE ONTOLOGY, C. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics,* 25(1), 25-29.

ASKEW, C., SELLAM, A., EPP, E., HOGUES, H., MULLICK, A., NANTEL, A. & WHITEWAY, M. 2009. Transcriptional Regulation of Carbohydrate Metabolism in the Human Pathogen *Candida albicans*. *PLoS Pathogens,* 5(1), e1000612.

BAKSHI, S. S. 2018. Erythematous Candidiasis. *Journal of Allergy and Clinical Immunology-in Practice,* 6(2), 646-646.

BANERJEE, M., THOMPSON, D. S., LAZZELL, A., CARLISLE, P. L., PIERCE, C., MONTEAGUDO, C., LOPEZ-RIBOT, J. L. & KADOSH, D. 2008. *UME6*, a novel filament-specific regulator of *Candida albicans* hyphal extension and virulence. *Molecular Biology of the Cell,* 19(4), 1354-1365.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* 315(5819), 1709-1712.

BARTKOWIAK, B. & GREENLEAF, A. L. 2011. Phosphorylation of RNAPII: To P-TEFb or not to P-TEFb? *Transcription,* 2(3), 115–119.

BASSO, V., D'ENFERT, C., ZNAIDI, S. & BACHELLIER-BASSI, S. 2019. From Genes to Networks: The Regulatory Circuitry Controlling *Candida albicans* Morphogenesis. *Current Topics in Microbiology and Immunology,* 422, 61–99.

BENSEN, E. S., MARTIN, S. J., LI, M. C., BERMAN, J. & DAVIS, D. A. 2004. Transcriptional profiling in *Candida albicans* reveals new adaptive responses to extracellular pH and functions for Rim101p. *Molecular Microbiology,* 54(5), 1335-1351.

BJORKLUND, S. & GUSTAFSSON, C. M. 2005. The yeast Mediator complex and its regulation. *Trends in Biochemical Sciences,* 30(5), 240-244.

BONGOMIN, F., GAGO, S., OLADELE, R. O. & DENNING, D. W. 2017. Global and Multi-National Prevalence of Fungal Diseases-Estimate Precision. *Journal of Fungi,* 3(4), 57.

BOTTCHER, B., POLLATH, C., STAIB, P., HUBE, B. & BRUNKE, S. 2016. *Candida* species Rewired Hyphae Developmental Programs for Chlamydospore Formation. *Frontiers in Microbiology,* 7, 1697.

BRAUN, B. R. & JOHNSON, A. D. 1997. Control of filament formation in *Candida albicans* by the transcriptional repressor *TUP1*. *Science,* 277(5322), 105-109.

BRAUN, B. R., KADOSH, D. & JOHNSON, A. D. 2001. *NRG1*, a repressor of filamentous growth in C.albicans, is down-regulated during filament induction. *EMBO Journal,* 20(17), 4753-4761.

BROUNS, S. J. J., JORE, M. M., LUNDGREN, M., WESTRA, E. R., SLIJKHUIS, R. J. H., SNIJDERS, A. P. L., DICKMAN, M. J., MAKAROVA, K. S., KOONIN, E. V. & VAN DER OOST, J. 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science,* 321(5891), 960-964.

BRUNO, V. M., WANG, Z., MARJANI, S. L., EUSKIRCHEN, G. M., MARTIN, J., SHERLOCK, G. & SNYDER, M. 2010. Comprehensive annotation of the transcriptome of the human fungal pathogen *Candida albicans* using RNA-seq. *Genome Research,* 20(10), 1451-1458.

BUIS, J., STONEHAM, T., SPEHALSKI, E. & FERGUSON, D. O. 2012. Mre11 regulates CtIP-dependent double-strand break repair by interaction with CDK2. *Nature Structural & Molecular Biology,* 19(2), 246-252.

CAMPANHA, N. H., NEPPELENBROEK, K. H., SPOLIDORIO, D. M. P., SPOLIDORIO, L. C. & PAVARINA, A. C. 2005. Phenotypic methods and commercial systems for the discrimination between *C. albicans* and *C. dubliniensis*. *Oral Diseases,* 11(6), 392-398.

CARLISLE, P. L., BANERJEE, M., LAZZELL, A., MONTEAGUDO, C., LOPEZ-RIBOT, J. L. & KADOSH, D. 2009. Expression levels of a filament-specific transcriptional regulator are sufficient to determine *Candida albicans* morphology and virulence. *Proceedings of the National Academy of Sciences of the United States of America,* 106(2), 599-604.

CASAMASSIMI, A. & NAPOLI, C. 2007. Mediator complexes and eukaryotic transcription regulation: An overview. *Biochimie,* 89(12), 1439-1446.

CHABRIER-ROSELLO, Y., GIESSELMAN, B. R., DE JESUS-ANDINO, F. J., FOSTER, T. H., MITRA, S. & HAIDARIS, C. G. 2010. Inhibition of electron transport chain assembly and function promotes photodynamic killing of *Candida*. *Journal of Photochemistry and Photobiology B-Biology,* 99(3), 117-125.

CHANDRA, J., KUHN, D. M., MUKHERJEE, P. K., HOYER, L. L., MCCORMICK, T. & GHANNOUM, M. A. 2001. Biofilm formation by the fungal pathogen *Candida albicans*: development, architecture, and drug resistance. *Journal of Bacteriology,* 183(18), 5385-9534.

CHATTAWAY, F. W., HOLMES, M. R. & BARLOW, A. J. E. 1968. Cell Wall Composition of Mycelial and Blastospore Forms of *Candida albicans*. *Journal of General Microbiology,* 51(3), 367-376.

CHEN, S. F., ZHOU, Y. Q., CHEN, Y. R. & GU, J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics,* 34(17), 884-890.

CHIBANA, H. & MAGEE, P. T. 2009. The enigma of the major repeat sequence of *Candida albicans*. *Future Microbiology,* 4(2), 171-179.

COHEN, J. 2019. Did CRISPR help—or harm—the first-ever gene-edited babies? *Science.* available: http://dx.doi.org/10.1126/science.aay9569.

COLLINS, S. R., MILLER, K. M., MAAS, N. L., ROGUEV, A., FILLINGHAM, J., CHU, C. S., SCHULDINER, M., GEBBIA, M., RECHT, J., SHALES, M., DING, H. M., XU, H., HAN, J. H., INGVARSDOTTIR, K., CHENG, B., ANDREWS, B., BOONE, C., BERGER, S. L., HIETER, P., ZHANG, Z. G., BROWN, G. W., INGLES, C. J., EMILI, A., ALLIS, C. D., TOCZYSKI, D. P., WEISSMAN, J. S., GREENBLATT, J. F. & KROGAN, N. J. 2007. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature,* 446(7137), 806-810.

CONG, L., RAN, F. A., COX, D., LIN, S., BARRETTO, R., HABIB, N., HSU, P. D., WU, X., JIANG, W., MARRAFFINI, L. A. & ZHANG, F. 2013. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science,* 339(6121), 819-823.

COSTANZO, M., BARYSHNIKOVA, A., BELLAY, J., KIM, Y., SPEAR, E. D., SEVIER, C. S., DING, H. M., KOH, J. L. Y., TOUFIGHI, K., MOSTAFAVI, S., PRINZ, J., ONGE, R. P. S., VANDERSLUIS, B., MAKHNEVYCH, T., VIZEACOUMAR, F. J., ALIZADEH, S., BAHR, S., BROST, R. L., CHEN, Y. Q., COKOL, M., DESHPANDE, R., LI, Z. J., LIN, Z. Y., LIANG, W. D., MARBACK, M., PAW, J., LUIS, B. J. S., SHUTERIQI, E., TONG, A. H. Y., VAN DYK, N., WALLACE, I. M., WHITNEY, J. A.,

WEIRAUCH, M. T., ZHONG, G. Q., ZHU, H. W., HOURY, W. A., BRUDNO, M., RAGIBIZADEH, S., PAPP, B., PAL, C., ROTH, F. P., GIAEVER, G., NISLOW, C., TROYANSKAYA, O. G., BUSSEY, H., BADER, G. D., GINGRAS, A. C., MORRIS, Q. D., KIM, P. M., KAISER, C. A., MYERS, C. L., ANDREWS, B. J. & BOONE, C. 2010. The Genetic Landscape of a Cell. *Science,* 327(5964), 425-431.

COSTE, A. T., KARABABA, M., ISCHER, F., BILLE, J. & SANGLARD, D. 2004. *TAC1*, transcriptional activator of *CDR* genes, is a new transcription factor involved in the regulation of *Candida albicans* ABC transporters *CDR1* and *CDR2*. *Eukaryotic Cell,* 3(6), 1639-1652.

DANECEK, P., BONFIELD, J. K., LIDDLE, J., MARSHALL, J., OHAN, V., POLLARD, M. O., WHITWHAM, A., KEANE, T., MCCARTHY, S. A., DAVIES, R. M. & LI, H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience,* 10(2), giab008.

DANTAS, A. D., DAY, A., IKEH, M., KOS, I., ACHAN, B. & QUINN, J. 2015. Oxidative Stress Responses in the Human Fungal Pathogen, *Candida albicans*. *Biomolecules,* 5(1), 142-165.

DE COSTER, W., D'HERT, S., SCHULTZ, D. T., CRUTS, M. & VAN BROECKHOVEN, C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics,* 34(15), 2666-2669.

DELTCHEVA, E., CHYLINSKI, K., SHARMA, C. M., GONZALES, K., CHAO, Y., PIRZADA, Z. A., ECKERT, M. R., VOGEL, J. & CHARPENTIER, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature,* 471(7340), 602-607.

DEO, P. N. & DESHMUKH, R. 2019. Oral microbiome: Unveiling the fundamentals. *Journal of Oral and Maxillofacial Pathology,* 23(1), 122-128.

DUNKER, C., POLKE, M., SCHULZE-RICHTER, B., SCHUBERT, K., RUDOLPHI, S., GRESSLER, A. E., PAWLIK, T., PRADA SALCEDO, J. P., NIEMIEC, M. J., SLESIONA-KUNZEL, S., SWIDERGALL, M., MARTIN, R., DANDEKAR, T. & JACOBSEN, I. D. 2021. Rapid proliferation due to better metabolic adaptation results in full virulence of a filament-deficient *Candida albicans* strain. *Nature Communications,* 12(1), 3899.

DUNN, M. J. & ANDERSON, M. Z. 2019. To Repeat or Not to Repeat: Repetitive Sequences Regulate Genome Stability in *Candida albicans*. *Genes,* 10(11), 866.

DUNN, M. J., KINNEY, G. M., WASHINGTON, P. M., BERMAN, J. & ANDERSON, M. Z. 2018. Functional diversification accompanies gene family expansion of *MED2* homologs in *Candida albicans*. *PLoS Genetics,* 4(4), e1007326.

EGUEZ, L., CHUNG, Y. S., KUCHIBHATLA, A., PAIDHUNGAT, M. & GARRETT, S. 2004. Yeast $Mn2^{+}$ transporter, Smf1p, is regulated by ubiquitin-dependent vacuolar protein sorting. *Genetics,* 167(1), 107-117.

ENJALBERT, B., MACCALLUM, D. M., ODDS, F. C. & BROWN, A. J. P. 2007. Niche-specific activation of the oxidative stress response by the pathogenic fungus *Candida albicans*. *Infection and Immunity,* 75(5), 2143-2151.

ENJALBERT, B., MORAN, G. P., VAUGHAN, C., YEOMANS, T., MACCALLUM, D. M., QUINN, J., COLEMAN, D. C., BROWN, A. J. P. & SULLIVAN, D. J. 2009. Genome-wide gene expression profiling and a forward genetic screen show that differential expression of the sodium ion transporter Ena21 contributes to the differential tolerance of *Candida albicans* and *Candida dubliniensis* to osmotic stress. *Molecular Microbiology,* 72(1), 216-228.

ENJALBERT, B., NANTEL, A. & WHITEWAY, M. 2003. Stress-induced gene expression in *Candida albicans*: Absence of a general stress response. *Molecular Biology of the Cell,* 14(4), 1460-1467.

FEKETE, A., EMRI, T., GYETVAI, A., GAZDAG, Z., PESTI, M., VARGA, Z., BALLA, J., CSERHATI, C., EMODY, L., GERGELY, L. & POCSI, I. 2007. Development of oxidative stress tolerance resulted in reduced ability to undergo morphologic transitions and decreased pathogenicity in a t-butylhydroperoxide-tolerant mutant of *Candida albicans*. *FEMS Yeast Research,* 7(6), 834-847.

FITZPATRICK, D. A., LOGUE, M. E., STAJICH, J. E. & BUTLER, G. 2006. A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evolutionary Biology,* 6, 99.

FLANAGAN, P. R., FLETCHER, J., BOYLE, H., SULEA, R., MORAN, G. P. & SULLIVAN, D. J. 2018. Expansion of the *TLO* gene family enhances the virulence of Candida species. *PLoS One,* 3(7), e0200852.

FOGARTY, N. M. E., MCCARTHY, A., SNIJDERS, K. E., POWELL, B. E., KUBIKOVA, N., LAKELEY, P. B., LEA, R., LDER, K. E., WAMAITHA, S. E., KIM, D., MACIULYTE, V., KLEINJUNG, J., KIM, J.-S.,

WELLS, D., VALLIER, L., ERTERO, A. B., URNER, J. M. A. T. & NIAKAN, K. K. 2017. Genome editing reveals a role for *OCT4* in human embryogenesis. *Nature,* 550(7674), 67-73.
FREESE, N. H., NORRIS, D. C. & LORAINE, A. E. 2016. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics,* 32(14), 2089-2095.
FU, M. S., DE SORDI, L. & MUHLSCHLEGEL, F. A. 2012. Functional Characterization of the Small Heat Shock Protein Hsp12p from *Candida albicans*. *PLoS One,* 7(8), e42894.
GARCIA-RUBIO, R., DE OLIVEIRA, H. C., RIVERA, J. & TREVIJANO-CONTADOR, N. 2020. The Fungal Cell Wall: *Candida, Cryptococcus*, and *Aspergillus* Species. *Frontiers in Microbiology,* 10, 2993.
GARNEAU, J. E., DUPUIS, M.-E., VILLION, M., ROMERO, D. A., BARRANGOU, R., BOYAVAL, P., FREMAUX, C., HORVATH, P., MAGADAN, A. H. & MOINEAU, S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature,* 468(7320), 67-71.
GILLUM, A. M., TSAY, E. Y. H. & KIRSCH, D. R. 1984. Isolation of the *Candida albicans* gene for orotidine-5′-phosphate decarboxylase by complementation of *S. cerevisiae URA3* and *Escherichia coli pyrF* mutations. *Molecular & General Genetics,* 198(2), 179-182.
GOFFRINI, P., FERRERO, I. & DONNINI, C. 2002. Respiration-dependent utilization of sugars in yeasts: a determinant role for sugar transporters. *Journal of Bacteriology,* 184(2), 427-432.
GONCALVES, B., FERREIRA, C., ALVES, C. T., HENRIQUES, M., AZEREDO, J. & SILVA, S. 2016. Vulvovaginal candidiasis: Epidemiology, microbiology and risk factors. *Critical Reviews in Microbiology,* 42(6), 905-927.
GOW, N. A. R. & YADAV, B. 2017. Microbe Profile: *Candida albicans*: a shape-changing, opportunistic pathogenic fungus of humans. *Microbiology,* 163(8), 1145-1147.
GOYARD, S., KNECHTLE, P., CHAUVEL, M., MALLET, A., PREVOST, M. C., PROUX, C., COPPEE, J. Y., SCHWARTZ, P., DROMER, F., PARK, H., FILLER, S. G., JANBON, G. & D'ENFERT, C. 2008. The Yak1 kinase is involved in the initiation and maintenance of hyphal growth in *Candida albicans*. *Molecular Biology of the Cell,* 19(5), 2251-2266.
HALLBERG, M., POLOZKOV, G. V., HU, G. Z., BEVE, J., GUSTAFSSON, C. M., RONNE, H. & BJORKLUND, S. 2004. Site-specific Srb10-dependent phosphorylation of the yeast mediator subunit Med2 regulates gene expression from the 2- μm plasmid. *Proceedings of the National Academy of Sciences of the United States of America,* 101(10), 3370-3375.
HARAN, J., BOYLE, H., HOKAMP, K., YEOMANS, T., LIU, Z., CHURCH, M., FLEMING, A. B., ANDERSON, M. Z., BERMAN, J., MYERS, L. C., SULLIVAN, D. J. & MORAN, G. P. 2014. Telomeric ORFs (*TLO*s) in *Candida* spp. Encode mediator subunits that regulate distinct virulence traits. *PLoS Genet,* 10(10), e1004658.
HARCUS, D., NANTEL, A., MARCIL, A., RIGBY, T. & WHITEWAY, M. 2004. Transcription profiling of cyclic AMP signaling in *Candida albicans*. *Molecular Biology of the Cell,* 15(10), 4490-4499.
HENRY, S. A., KOHLWEIN, S. D. & CARMAN, G. M. 2012. Metabolism and regulation of glycerolipids in the yeast *Saccharomyces cerevisiae*. *Genetics,* 190(2), 317-349.
HIRAKAWA, M. P., MARTINEZ, D. A., SAKTHIKUMAR, S., ANDERSON, M. Z., BERLIN, A., GUJJA, S., ZENG, Q., ZISSON, E., WANG, J. M., GREENBERG, J. M., BERMAN, J., BENNETT, R. J. & CUOMO, C. A. 2015. Genetic and phenotypic intra-species variation in *Candida albicans*. *Genome Research,* 25(3), 413-425.
HO, J. W. K., BISHOP, E., KARCHENKO, P. V., NEGRE, N., WHITE, K. P. & PARK, P. J. 2011. ChIP-chip versus ChIP-seq: Lessons for experimental design and data analysis. *BMC Genomics,* 12, 134.
HOSSAIN, S., VERI, A. O. & COWEN, L. E. 2020. The Proteasome Governs Fungal Morphogenesis via Functional Connections with Hsp90 and cAMP-Protein Kinase A Signaling. *Mbio,* 11(2), e00290-20.
HSU, P. D., LANDER, E. S. & ZHANG, F. 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell,* 157(6), 1262-1278.
HUTTENRACH, F., POLLOK-KOPP, B. & OPPERMANN, M. 2005. G protein-coupled receptor kinases promote phosphorylation and beta-arrestin-mediated internalization of *CCR5* homo- and hetero-oligomers. *Journal of Biological Chemistry,* 280(45), 37503-37515.
ISHINO, Y., SHINAGAWA, H., MAKINO, K., AMEMURA, M. & NAKATA, A. 1987. Nucleotide-sequence of the *iap* gene, responsible for alkaline-phosphatase isozyme conversion in

Escherichia coli, and identification of the gene-product. *Journal of Bacteriology,* 169(12), 5429-5433.

JABRA-RIZK, M. A., KONG, E. F., TSUI, C., HONG NGUYEN, M., CLANCY, C. J., FIDEL, P. L., JR. & NOVERR, M. 2016. *Candida albicans* Pathogenesis: Fitting within the Host-Microbe Damage Response Framework. *Infection and Immunity,* 84(10), 2724-2739.

JACKSON, A. P., GAMBLE, J. A., YEOMANS, T., MORAN, G. P., SAUNDERS, D., HARRIS, D., ASLETT, M., BARRELL, J. F., BUTLER, G., CITIULO, F., COLEMAN, D. C., DE GROOT, P. W., GOODWIN, T. J., QUAIL, M. A., MCQUILLAN, J., MUNRO, C. A., PAIN, A., POULTER, R. T., RAJANDREAM, M. A., RENAULD, H., SPIERING, M. J., TIVEY, A., GOW, N. A., BARRELL, B., SULLIVAN, D. J. & BERRIMAN, M. 2009. Comparative genomics of the fungal pathogens *Candida dubliniensis* and *Candida albicans*. *Genome Research,* 19(12), 2231-2244.

JAMES, J., FIJI, N., ROY, D., ANDREW, M. G. D., SHIHABUDEEN, M. S., CHATTOPADHYAY, D. & THIRUMURUGAN, K. 2015. A rapid method to assess reactive oxygen species in yeast using H_2DCF-DA. *Analytical Methods,* 7(20), 8572-8575.

JANSEN, R., VAN EMBDEN, J. D. A., GAASTRA, W. & SCHOULS, L. M. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular Microbiology,* 43(6), 1565-1575.

JINEK, M., CHYLINSKI, K., FONFARA, I., HAUER, M., DOUDNA, J. A. & CHARPENTIER, E. 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science,* 337(6096), 816-821.

JONES, T., FEDERSPIEL, N. A., CHIBANA, H., DUNGAN, J., KALMAN, S., MAGEE, B. B., NEWPORT, G., THORSTENSON, Y. R., AGABIAN, N., MAGEE, P. T., DAVIS, R. W. & SCHERER, S. 2004. The diploid genome sequence of *Candida albicans*. *Proceedings of the National Academy of Sciences of the United States of America,* 101(19), 7329-7334.

KALORITI, D., TILLMANN, A., COOK, E., JACOBSEN, M., YOU, T., LENARDON, M., AMES, L., BARAHONA, M., CHANDRASEKARAN, K., COGHILL, G., GOODMAN, D., GOW, N. A. R., GREBOGI, C., HO, H. L., INGRAM, P., MCDONAGH, A., DE MOURA, A. P. S., PANG, W., PUTTNAM, M., RADMANESHFAR, E., ROMANO, M. C., SILK, D., STARK, J., STUMPF, M., THIEL, M., THORNE, T., USHER, J., YIN, Z. K., HAYNES, K. & BROWN, A. J. P. 2012. Combinatorial stresses kill pathogenic *Candida* species. *Medical Mycology,* 50(7), 699-709.

KARABABA, M., VALENTINO, E., PARDINI, G., COSTE, A. T., BILLE, J. & SANGLARD, D. 2006. *CRZ1*, a target of the calcineurin pathway in *Candida albicans*. *Molecular Microbiology,* 59(5), 1429-1451.

KELLNER, M. J., KOOB, J. G., GOOTENBERG, J. S., ABUDAYYEH, O. O. & ZHANG, F. 2019. SHERLOCK: nucleic acid detection with CRISPR nucleases. *Nature Protocols,* 14(10), 2986-3012.

KENT, N. A., ADAMS, S., MOORHOUSE, A. & PASZKIEWICZ, K. 2011. Chromatin particle spectrum analysis: a method for comparative chromatin structure analysis using paired-end mode next-generation DNA sequencing. *Nucleic Acids Research,* 39(5), e26.

KENT, W. J., ZWEIG, A. S., BARBER, G., HINRICHS, A. S. & KAROLCHIK, D. 2010. BigWig and BigBed: enabling browsing of large distributed datasets. *Bioinformatics,* 26(17), 2204-2207.

KESSLER, G. & NICKERSON, W. J. 1959. Glucomannan-protein complexes from cell walls of yeast. *Journal of Biological Chemistry,* 234, 2281-2285.

KNIGHT, S. A. B., TAMAI, K. T., KOSMAN, D. J. & THIELE, D. J. 1994. Identification and analysis of a *Saccharomyces cerevisiae* copper homeostasis gene encoding a homeodomain protein. *Molecular and Cellular Biology,* 14(12), 7792-7804.

KOONIN, E. V. & MAKAROVA, K. S. 2019. Origins and evolution of CRISPR-Cas systems. *Philosophical Transactions of the Royal Society B-Biological Sciences,* 374(1772), 20180087.

KULLBERG, B. J. & ARENDRUP, M. C. 2015. Invasive Candidiasis. *New England Journal of Medicine,* 373(15), 1445-1456.

LAGREE, K., WOOLFORD, C. A., HUANG, M. N. Y., MAY, G., MCMANUS, C. J., SOLIS, N. V., FILLER, S. G. & MITCHELL, A. P. 2020. Roles of *Candida albicans* Mig1 and Mig2 in glucose repression, pathogenicity traits, and *SNF1* essentiality. *PLoS Genetics,* 16(1), e1008582.

LAMOTH, F., LOCKHART, S. R., BERKOW, E. L. & CALANDRA, T. 2018. Changes in the epidemiological landscape of invasive candidiasis. *Journal of Antimicrobial Chemotherapy,* 73(suppl_1), i4-i13.

LANDT, S. G., MARINOV, G. K., KUNDAJE, A., KHERADPOUR, P., PAULI, F., BATZOGLOU, S., BERNSTEIN, B. E., BICKEL, P., BROWN, J. B., CAYTING, P., CHEN, Y. W., DESALVO, G.,

EPSTEIN, C., FISHER-AYLOR, K. I., EUSKIRCHEN, G., GERSTEIN, M., GERTZ, J., HARTEMINK, A. J., HOFFMAN, M. M., IYER, V. R., JUNG, Y. L., KARMAKAR, S., KELLIS, M., KHARCHENKO, P. V., LI, Q. H., LIU, T., LIU, X. S., MA, L. J., MILOSAVLJEVIC, A., MYERS, R. M., PARK, P. J., PAZIN, M. J., PERRY, M. D., RAHA, D., REDDY, T. E., ROZOWSKY, J., SHORESH, N., SIDOW, A., SLATTERY, M., STAMATOYANNOPOULOS, J. A., TOLSTORUKOV, M. Y., WHITE, K. P., XI, S., FARNHAM, P. J., LIEB, J. D., WOLD, B. J. & SNYDER, M. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research,* 22(9), 1813-1831.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods,* 9(4), 357-U54.

LEDFORD, H. & CALLAWAY, E. 2020. Pioneers of revolutionary CRISPR gene editing win chemistry Nobel. *Nature,* 586(7829), 346-347.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA, P. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25(16), 2078-2079.

LI, H. W. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv: Genomics*, available: https://arxiv.org/abs/1303.3997

LIANG, P., XU, Y., ZHANG, X., DING, C., HUANG, R., ZHANG, Z., LV, J., XIE, X., CHEN, Y., LI, Y., SUN, Y., BAI, Y., SONGYANG, Z., MA, W., ZHOU, C. & HUANG, J. 2015. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein & Cell,* 6(5), 363-372.

LIU, T. T., LEE, R. E. B., BARKER, K. S., LEE, R. E., WEI, L., HOMAYOUNI, R. & ROGERS, P. D. 2005. Genome-wide expression profiling of the response to azole, polyene, echinocandin, and pyrimidine antifungal agents in *Candida albicans*. *Antimicrobial Agents and Chemotherapy,* 49(6), 2226-2236.

LIU, T. T., ZNAIDI, S., BARKER, K. S., XU, L., HOMAYOUNI, R., SAIDANE, S., MORSCHHAUSER, J., NANTEL, A., RAYMOND, M. & ROGERS, P. D. 2007. Genome-wide expression and location analyses of the *Candida albicans* Tac1p regulon. *Eukaryotic Cell,* 6(11), 2122-2138.

LIU, Y. & FILLER, S. G. 2011. *Candida albicans* Als3, a Multifunctional Adhesin and Invasin. *Eukaryotic Cell,* 10(2), 168-173.

LIU, Z., MORAN, G. P., SULLIVAN, D. J., MACCALLUM, D. M. & MYERS, L. C. 2016. Amplification of *TLO* Mediator Subunit Genes Facilitate Filamentous Growth in *Candida* Spp. *PLoS Genetics,* 12(10), e1006373.

LIU, Z. & MYERS, L. C. 2015. Fungal Mediator Tail Subunits Contain Classical Transcriptional Activation Domains. *Molecular and Cellular Biology,* 35(8), 1363-1375.

LOHSE, M. B. & JOHNSON, A. D. 2009. White-opaque switching in *Candida albicans*. *Current Opinion in Microbiology,* 12(6), 650-654.

LORENZ, M. C., BENDER, J. A. & FINK, G. R. 2004. Transcriptional response of *Candida albicans* upon internalization by macrophages. *Eukaryotic Cell,* 3(5), 1076-1087.

LORETO, E. S., SCHEID, L. A., NOGUEIRA, C. W., ZENI, G., SANTURIO, J. M. & ALVES, S. H. 2010. *Candida dubliniensis*: Epidemiology and Phenotypic Methods for Identification. *Mycopathologia,* 169(6), 431-443.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology,* 15, 550.

MACPHERSON, S., AKACHE, B., WEBER, S., DE DEKEN, X., RAYMOND, M. & TURCOTTE, B. 2005. *Candida albicans* zinc cluster protein Upc2p confers resistance to antifungal drugs and is an activator of ergosterol biosynthetic genes. *Antimicrobial Agents and Chemotherapy,* 49(5), 1745-1752.

MAGEE, B., SANCHEZ, M. D., SAUNDERS, D., HARRIS, D., BERRIMAN, M., MAGEE P. T. 2008. Extensive chromosome rearrangements distinguish the karyotype of the hypovirulent species *Candida dubliniensis* from the virulent *Candida albicans. Fungal Genetics and Biology,* 45(3), 338–350.

MALI, P., YANG, L., ESVELT, K. M., AACH, J., GUELL, M., DICARLO, J. E., NORVILLE, J. E. & CHURCH, G. M. 2013. RNA-Guided Human Genome Engineering via Cas9. *Science,* 339(6121), 823-826.

MARTINEZ, J. P., BLANES, R., CASANOVA, M., VALENTIN, E., MURGUI, A. & DOMINGUEZ, A. 2016. Null mutants of *Candida albicans* for cell-wall-related genes form fragile biofilms that display an almost identical extracellular matrix proteome. *FEMS Yeast Research,* 16(7), fow075.

MAYER, F. L., WILSON, D. & HUBE, B. 2013. *Candida albicans* pathogenicity mechanisms. *Virulence,* 4(2), 119-28.
MCCOOL, L., MAI, H., ESSMANN, M. & LARSEN, B. 2008. Tetracycline effects on *Candida albicans* virulence factors. *Infectious Diseases in Obstetrics and Gynecology,* 2008, 493508-493508.
MCMANUS, B. A. & COLEMAN, D. C. 2014. Molecular epidemiology, phylogeny and evolution of *Candida albicans*. *Infection Genetics and Evolution,* 21, 166-178.
MOJICA, F. J. M., DIEZ-VILLASENOR, C., GARCIA-MARTINEZ, J. & SORIA, E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution,* 60(2), 174-182.
MOJICA, F. J. M., DIEZ-VILLASENOR, C., SORIA, E. & JUEZ, G. 2000. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, bacteria and mitochondria. *Molecular Microbiology,* 36(1), 244-246.
MORAN, G. P., ANDERSON, M. Z., MYERS, L. C. & SULLIVAN, D. J. 2019. Role of Mediator in virulence and antifungal drug resistance in pathogenic fungi. *Current Genetics,* 65(3), 621-630.
MORAN, G. P., COLEMAN, D. C. & SULLIVAN, D. J. 2012. *Candida albicans* versus *Candida dubliniensis*: Why Is *C. albicans* More Pathogenic? *International Journal of Microbiology,* 2012, 205921.
MORAN, G. P., MACCALLUM, D. M., SPIERING, M. J., COLEMAN, D. C. & SULLIVAN, D. J. 2007. Differential regulation of the transcriptional repressor *NRG1* accounts for altered host-cell interactions in *Candida albicans* and *Candida dubliniensis*. *Molecular Microbiology,* 66(4), 915-929.
MORSCHHAUSER, J., RUHNKE, M., MICHEL, S. & HACKER, J. 1999. Identification of CARE-2-negative *Candida albicans* isolates as *Candida dubliniensis*. *Mycoses,* 42(1-2), 29-32.
MURAD, A. M. A., LENG, P., STRAFFON, M., WISHART, J., MACASKILL, S., MACCALLUM, D., SCHNELL, N., TALIBI, D., MARECHAL, D., TEKAIA, F., D'ENFERT, C., GAILLARDIN, C., ODDS, F. C. & BROWN, A. J. P. 2001. *NRG1* represses yeast-hypha morphogenesis and hypha-specific gene expression in *Candida albicans*. *EMBO Journal,* 20(17), 4742-4752.
MUZYKA, B. C. & EPIFANIO, R. N. 2013. Update on oral fungal infections. *Dental Clinics of North America,* 57(4), 561-581.
NAKANISHI, H., SUDA, Y. & NEIMAN, A. M. 2007. Erv14 family cargo receptors are necessary for ER exit during sporulation in *Saccharomyces cerevisiae*. *Journal of Cell Science,* 120(Pt 5), 908-916.
NAVARRO-GARCIA, F., EISMAN, B., FIUZA, S. M., NOMBELA, C. & PLA, J. 2005. The MAP kinase Mkc1p is activated under different stress conditions in *Candida albicans*. *Microbiology (Reading),* 151(Pt 8), 2737-2749.
NGUYEN, N., QUAIL, M. M. F. & HERNDAY, A. D. 2017. An Efficient, Rapid, and Recyclable System for CRISPR-Mediated Genome Editing in *Candida albicans*. *Msphere,* 2(2), e00149-17.
NOBILE, C. J., BRUNO, V. M., RICHARD, M. L., DAVIS, D. A. & MITCHELL, A. P. 2003. Genetic control of chlamydospore formation in *Candida albicans*. *Microbiology (Reading),* 149(Pt 12), 3629-3637.
NOBILE, C. J., FOX, E. P., NETT, J. E., SORRELLS, T. R., MITROVICH, Q. M., HERNDAY, A. D., TUCH, B. B., ANDES, D. R. & JOHNSON, A. D. 2012. A Recently Evolved Transcriptional Network Controls Biofilm Development in *Candida albicans*. *Cell,* 148(1-2), 126-138.
NOBILE, C. J. & JOHNSON, A. D. 2015. *Candida albicans* Biofilms and Human Disease. *Annual Review of Microbiology,* 69, 71-92.
NOBLE, S. M., GIANETTI, B. A. & WITCHLEY, J. N. 2017. *Candida albicans* cell-type switching and functional plasticity in the mammalian host. *Nature Reviews Microbiology,* 15(2), 96-108.
NORMILE, D. 2019. Chinese scientist who produced genetically altered babies sentenced to 3 years in jail. *Science.* available: http://dx.doi.org/10.1126/science.aba7347.
O'CONNOR, L., CAPLICE, N., COLEMAN, D. C., SULLIVAN, D. J. & MORAN, G. P. 2010. Differential Filamentation of *Candida albicans* and *Candida dubliniensis* Is Governed by Nutrient Regulation of *UME6* Expression. *Eukaryotic Cell,* 9(9), 1383-1397.
ODDS, F. C., BROWN, A. J. & GOW, N. A. 2004. *Candida albicans* genome sequence: a platform for genomics in the absence of genetics. *Genome Biology,* 5(7), 230.
OHTA, K. & YOSHIMURA, H. 2020. Leukoplakia of the tongue. *Cleveland Clinic Journal of Medicine,* 87(3), 133-134.

ORPHANIDES, G., LAGRANGE, T. & REINBERG, D. 1996. The general transcription factors of RNA polymerase II. *Genes & Development,* 10(21), 2657-2683.

PADMANABHAN, S., THAKUR, J., SIDDHARTHAN, R., SANYAL, K. 2008. Rapid evolution of Cse4p-rich centromeric DNA sequences in closely related pathogenic yeasts, *Candida albicans* and *Candida dubliniensis*. *Proceedings of the National Academy of Sciences of the United States of America*, 105(50), 19797-19802.

PALIGE, K., LINDE, J., MARTIN, R., BOTTCHER, B., CITIULO, F., SULLIVAN, D. J., WEBER, J., STAIB, C., RUPP, S., HUBE, B., MORSCHHAUSER, J. & STAIB, P. 2013. Global Transcriptome Sequencing Identifies Chlamydospore Specific Markers in *Candida albicans* and *Candida dubliniensis*. *PLoS One,* 8(4), e61940.

PAN, X. W., YE, P., YUAN, D. S., WANG, X. L., BADER, J. S. & BOEKE, J. D. 2006. A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell,* 124(5), 1069-1081.

PAPPAS, P. C., LIONAKIS, M. S., ARENDRUP, M. C., OSTROSKY-ZEICHNER, L. & KULLBERG, B. J. 2018. Invasive candidiasis. *Nature Reviews Disease Primers,* 4, 18026.

PARK, Y. N. & MORSCHHAUSER, J. 2005. Tetracycline-inducible gene expression and gene deletion in *Candida albicans*. *Eukaryotic Cell,* 4(8), 1328-1342.

PATIL, S., MAJUMDAR, B., SARODE, S. C., SARODE, G. S. & AWAN, K. H. 2018. Oropharyngeal Candidosis in HIV-Infected Patients - An Update. *Frontiers in Microbiology,* 9, 980.

PEREIRA, T. C., DE BARROS, P. P., FUGISAKI, L. R. D., ROSSONI, R. D., RIBEIRO, F. D., DE MENEZES, R. T., JUNQUEIRA, J. C. & SCORZONI, L. 2018. Recent Advances in the Use of *Galleria mellonella* Model to Study Immune Responses against Human Pathogens. *Journal of Fungi,* 4(4), 128.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26(6), 841-842.

RAM, A. F. J. & KLIS, F. M. 2006. Identification of fungal cell wall mutants using susceptibility assays based on Calcofluor white and Congo red. *Nature Protocols,* 1(5), 2253-2256.

RAO, M. S., VAN VLEET, T. R., CIURLIONIS, R., BUCK, W. R., MITTELSTADT, S. W., BLOMME, E. A. G. & LIGUORI, M. J. 2019. Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. *Frontiers in Genetics,* 9, 636.

REARDON, S. 2016. First CRISPR clinical trial gets green light from US panel. *Nature*, available: https://doi.org/10.1038/nature.2016.20137

REUSS, O., VIK, A., KOLTER, R. & MORSCHHAUSER, J. 2004. The *SAT1* flipper, an optimized tool for gene disruption in *Candida albicans*. *Gene,* 341, 119-127.

ROBINSON, J. T., THORVALDSDOTTIR, H., WINCKLER, W., GUTTMAN, M., LANDER, E. S., GETZ, G. & MESIROV, J. P. 2011. Integrative genomics viewer. *Nature Biotechnology,* 29(1), 24-26.

ROMAN, E., COMAN, I., PRIETO, D., ALONSO-MONGE, R. & PLA, J. 2019. Implementation of a CRISPR-Based System for Gene Regulation in *Candida albicans*. *Msphere,* 4(1), :e00001-19.

ROMAN, E., NOMBELA, C. & PLA, J. 2005. The Sho1 adaptor protein links oxidative stress to morphogenesis and cell wall biosynthesis in the fungal pathogen *Candida albicans*. *Molecular and Cellular Biology,* 25(23), 10611-10627.

ROSATI, D., BRUNO, M., JAEGER, M., TEN OEVER, J. & NETEA, M. G. 2020. Recurrent Vulvovaginal Candidiasis: An Immunological Perspective. *Microorganisms,* 8(2), 144.

ROSENTUL, D. C., DELSING, C. E., JAEGER, M., PLANTINGA, T. S., OOSTING, M., COSTANTINI, I., VENSELAAR, H., JOOSTEN, L. A. B., VAN DER MEER, J. W. M., DUPONT, B., KULLBERG, B.-J., SOBEL, J. D. & NETEA, M. G. 2014. Gene polymorphisms in pattern recognition receptors and susceptibility to idiopathic recurrent vulvovaginal candidiasis. *Frontiers in Microbiology,* 5, 483.

RUPNIK, A., GRENON, M. & LOWNDES, N. 2008. The MRN complex. *Current Biology,* 18(11), PR455-R457.

SALUCCI, V., SCARITO, A., AURISICCHIO, L., LAMARTINA, S., NICOLAUS, G., GIAMPAOLI, S., GONZALEZ-PAZ, O., TONIATTI, C., BUJARD, H., HILLEN, W., CILIBERTO, G. & PALOMBO, F. 2002. Tight control of gene expression by a helper-dependent adenovirus vector carrying the rtTA2(s)-M2 tetracycline transactivator and repressor system. *Gene Therapy,* 9(21), 1415-1421.

SARDI, J. C. O., SCORZONI, L., BERNARDI, T., FUSCO-ALMEIDA, A. M. & GIANNINI, M. 2013. *Candida* species: current epidemiology, pathogenicity, biofilm formation, natural

antifungal products and new therapeutic options. *Journal of Medical Microbiology,* 62(Pt 1), 10-24.
SCHIER, A. C. & TAATJES, D. J. 2020. Structure and mechanism of the RNA polymerase II transcription machinery. *Genes & Development,* 34(7-8), 465-488.
SCHMIDT, D., WILSON, M. D., SPYROU, C., BROWN, G. D., HADFIELD, J. & ODOM, D. T. 2009. ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions. *Methods,* 48(3), 240-248.
SCHMITTGEN, T. D. & LIVAK, K. J. 2008. Analyzing real-time PCR data by the comparative C-T method. *Nature Protocols,* 3(6), 1101-1108.
SELMECKI, A., GERAMI-NEJAD, M., PAULSON, C., FORCHE, A. & BERMAN, J. 2008. An isochromosome confers drug resistance in vivo by amplification of two genes, *ERG11* and *TAC1*. *Molecular Microbiology,* 68(3), 624-641.
SHIIO, Y., EISENMAN, R. N., YI, E. C., DONOHOE, S., GOODLETT, D. R. & AEBERSOLD, R. 2003. Quantitative proteomic analysis of chromatin-associated factors. *Journal of the American Society for Mass Spectrometry,* 14(7), 696-703.
SHORT, J. M., FERNANDEZ, J. M., SORGE, J. A. & HUSE, W. D. 1988. Lambda ZAP - A Bacteriophage Lambda Expression Vector with *in vivo* Excision Properties. *Nucleic Acids Research,* 16(15), 7583-7600.
SKRZYPEK, M. S., BINKLEY, J., BINKLEY, G., MIYASATO, S. R., SIMISON, M. & SHERLOCK, G. 2017. The *Candida* Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Research,* 45(Database issue), D592-D596.
SONNEBORN, A., BOCKMUHL, D. P. & ERNST, J. F. 1999. Chlamydospore formation in *Candida albicans* requires the Efg1p morphogenetic regulator. *Infection and Immunity,* 67(10), 5514-5517.
SOUTOURINA, J. 2018. Transcription regulation by the Mediator complex. *Nature Reviews Molecular Cell Biology,* 19(4), 262-274.
STADTMAUER, E. A., FRAIETTA, J. A., DAVIS, M. M., COHEN, A. D., WEBER, K. L., LANCASTER, E., MANGAN, P. A., KULIKOVSKAYA, I., GUPTA, M., CHEN, F., TIAN, L., GONZALEZ, V. E., XU, J., JUNG, I.-Y., MELENHORST, J. J., PLESA, G., SHEA, J., MATLAWSKI, T., CERVINI, A., GAYMON, A. L., DESJARDINS, S., LAMONTAGNE, A., SALAS-MCKEE, J., FESNAK, A., SIEGEL, D. L., LEVINE, B. L., JADLOWSKY, J. K., YOUNG, R. M., CHEW, A., HWANG, W.-T., HEXNER, E. O., CARRENO, B. M., NOBLES, C. L., BUSHMAN, F. D., PARKER, K. R., QI, Y., SATPATHY, A. T., CHANG, H. Y., ZHAO, Y., LACEY, S. F. & JUNE, C. H. 2020. CRISPR-engineered T cells in patients with refractory cancer. *Science,* 367(6481), eaba7365.
STAIB, P., MORAN, G. P., SULLIVAN, D. J., COLEMAN, D. C. & MORSCHHAUSER, J. 2001. Isogenic strain construction and gene targeting in *Candida dubliniensis*. *Journal of Bacteriology,* 183(9), 2859-2865.
STAIB, P. & MORSCHHAEUSER, J. 2007. Chlamydospore formation in *Candida albicans* and *Candida dubliniensis* - an enigmatic developmental programme. *Mycoses,* 50(1), 1-12.
SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S. & MESIROV, J. P. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America,* 102(43), 15545-15550.
SUDBERY, P., GOW, N. & BERMAN, J. 2004. The distinct morphogenic states of *Candida albicans*. *Trends in Microbiology,* 12(7), 317-324.
SUDBERY, P. E. 2011. Growth of *Candida albicans* hyphae. *Nature Reviews Microbiology,* 9(1), 737-748.
SULLIVAN, D. & COLEMAN, D. 1998. *Candida dubliniensis*: Characteristics and identification. *Journal of Clinical Microbiology,* 36(2), 329-334.
SULLIVAN, D. J., BERMAN, J., MYERS, L. C. & MORAN, G. P. 2015. Telomeric ORFS in *Candida albicans*: does mediator tail wag the yeast? *PLoS Pathogens,* 11(2), e1004614.
SULLIVAN, D. J., MORAN, G. P. & COLEMAN, D. C. 2005. *Candida dubliniensis*: Ten years on. *FEMS Microbiology Letters,* 253(1), 9-17.
SULLIVAN, D. J., WESTERNENG, T. J., HAYNES, K. A., BENNETT, D. E. & COLEMAN, D. C. 1995. *Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. *Microbiology,* 141 (Pt 7), 1507-1521.

TAKAGI, J., SINGH-BABAK, S. D., LOHSE, M. B., DALAL, C. K. & JOHNSON, A. D. 2019. *Candida albicans* white and opaque cells exhibit distinct spectra of organ colonization in mouse models of infection. *PLoS One,* 14(6), e0218037.

TALAPKO, J., JUZBASIC, M., MATIJEVIC, T., PUSTIJANAC, E., BEKIC, S., KOTRIS, I. & SKRLEC, I. 2021. *Candida albicans*-The Virulence Factors and Clinical Manifestations of Infection. *Journal of Fungi,* 7(2), 79.

TALIBI, D. & RAYMOND, M. 1999. Isolation of a putative *Candida albicans* transcriptional regulator involved in pleiotropic drug resistance by functional complementation of a pdr1 pdr3 mutation in *Saccharomyces cerevisiae*. *Journal of Bacteriology,* 181(1), 231-240.

TAN, K. & WONG, K. H. 2019. RNA polymerase II ChIP-seq-a powerful and highly affordable method for studying fungal genomics and physiology. *Biophysical Reviews,* 11(1), 79-82.

TEBBJI, F., CHEN, Y., RICHARD ALBERT, J., GUNSALUS, K. T. W., KUMAMOTO, C. A., NANTEL, A., SELLAM, A. & WHITEWAY, M. 2014. A Functional Portrait of Med7 and the Mediator Complex in *Candida albicans*. *PLoS Genetics,* 10(11), e1004770.

TEBBJI, F., CHEN, Y. L., SELLAM, A. & WHITEWAY, M. 2017. The Genomic Landscape of the Fungus-Specific SWI/SNF Complex Subunit, Snf6, in *Candida albicans*. *Msphere,* 2(6), e00497-17.

THOMPSON, D. S., CARLISLE, P. L. & KADOSH, D. 2011. Coevolution of Morphology and Virulence in *Candida* Species. *Eukaryotic Cell,* 10(9), 1173-1182.

TSUTSUI, T., FUKASAWA, R., SHINMYOUZU, K., NAKAGAWA, R., TOBE, K., TANAKA, A. & OHKUMA, Y. 2013. Mediator Complex Recruits Epigenetic Regulators via Its Two Cyclin-dependent Kinase Subunits to Repress Transcription of Immune Response Genes. *Journal of Biological Chemistry,* 288(29), 20955-20965.

TURNER, S. A. & BUTLER, G. 2014. The *Candida* Pathogenic Species Complex. *Cold Spring Harbor Perspectives in Medicine,* 4(9), a019778.

UPPULURI, P., ZALDIVAR, M. A., ANDERSON, M. Z., DUNN, M. J., BERMAN, J., RIBOT, J. L. L. & KOHLER, J. R. 2018. *Candida albicans* Dispersed Cells Are Developmentally Distinct from Biofilm and Planktonic Cells. *Mbio,* 9(4), e01338-18.

UTHAYAKUMAR, D., SHARMA, J., WENSING, L. & SHAPIRO, R. S. 2020. CRISPR-Based Genetic Manipulation of *Candida* Species: Historical Perspectives and Current Approaches. *Frontiers in Genome Editing,* 2, 606281-606281.

UWAMAHORO, N., QU, Y., JELICIC, B., LO, T. L., BEAUREPAIRE, C., BANTUN, F., QUENAULT, T., BOAG, P. R., RAMM, G., CALLAGHAN, J., BEILHARZ, T. H., NANTEL, A., PELEG, A. Y. & TRAVEN, A. 2012. The Functions of Mediator in *Candida albicans* Support a Role in Shaping Species-Specific Gene Expression. *PLoS Genetics,* 8(4), e1002613.

VAN HET HOOG, M., RAST, T. J., MARTCHENKO, M., GRINDLE, S., DIGNARD, D., HOGUES, H., CUOMO, C., BERRIMAN, M., SCHERER, S., MAGEE, B. B., WHITEWAY, M., CHIBANA, H., NANTEL, A. & MAGEE, P. T. 2007. Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. *Genome Biology,* 8, R52.

VASICEK, E. M., BERKOW, E. L., FLOWERS, S. A., BARKER, K. S. & ROGERS, P. D. 2014. *UPC2* Is Universally Essential for Azole Antifungal Resistance in *Candida albicans*. *Eukaryotic Cell,* 13(7), 933-946.

VILA, T., SULTAN, A. S., MONTELONGO-JAUREGUI, D. & JABRA-RIZK, M. A. 2020. Oral Candidiasis: A Disease of Opportunity. *Journal of Fungi,* 6(1), 15.

VYAS, V. K., BARRASA, M. I. & FINK, G. R. 2015. A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families. *Science Advances,* 1(3), e1500248.

WAL, M. & PUGH, B. F. 2012. Genome-Wide Mapping of Nucleosome Positions in Yeast Using High-Resolution MNase ChIP-Seq. *Methods in Enzymology*, 513, 233–250.

WHITMORE, S. E. & LAMONT, R. J. 2014. Oral Bacteria and Cancer. *PLoS Pathogens,* 10(3), e1003933.

WICK, R. R., JUDD, L. M., GORRIE, C. L. & HOLT, K. E. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microbial Genomics,* 3(10), e000132.

WICKHAM, H. 2009. ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag*, ISBN: 978-3-319-24277-4.

WIEDENHEFT, B., STERNBERG, S. H. & DOUDNA, J. A. 2012. RNA-guided genetic silencing systems in bacteria and archaea. *Nature,* 482(7385), 331-338.

WILLEMS, H. M. E., AHMED, S. S., LIU, J. Y., XU, Z. B. & PETERS, B. M. 2020. Vulvovaginal Candidiasis: A Current Understanding and Burning Questions. *Journal of Fungi,* 6(1), 27.

WILLGER, S. D., LIU, Z., OLARTE, R. A., ADAMO, M. E., STAJICH, J. E., MYERS, L. C., KETTENBACH, A. N. & HOGAN, D. A. 2015. Analysis of the *Candida albicans* Phosphoproteome. *Eukaryotic Cell,* 14(5), 474-485.

WOOLFORD, J. L. & BASERGA, S. J. 2013. Ribosome Biogenesis in the Yeast *Saccharomyces cerevisiae*. *Genetics,* 195(3), 643-681.

WU, J. Q. & SNYDER, M. 2008. RNA polymerase II stalling: loading at the start prepares genes for a sprint. *Genome Biology,* 9(5), 220.

WU, Y. Q., WU, M. Y., WANG, Y. Y., CHEN, Y. S., GAO, J. & YING, C. M. 2018. *ERG11* couples oxidative stress adaptation, hyphal elongation and virulence in *Candida albicans*. *FEMS Yeast Research,* 18(7).

XIE, J., TAO, L., NOBILE, C. J., TONG, Y. J., GUAN, G. B., SUN, Y., CAO, C. J., HERNDAY, A. D., JOHNSON, A. D., ZHANG, L. X., BAI, F. Y. & HUANG, G. H. 2013. White-Opaque Switching in Natural *MTL*a/α Isolates of *Candida albicans*: Evolutionary Implications for Roles in Host Adaptation, Pathogenesis, and Sex. *PLoS Biology,* 11(3), e1001525.

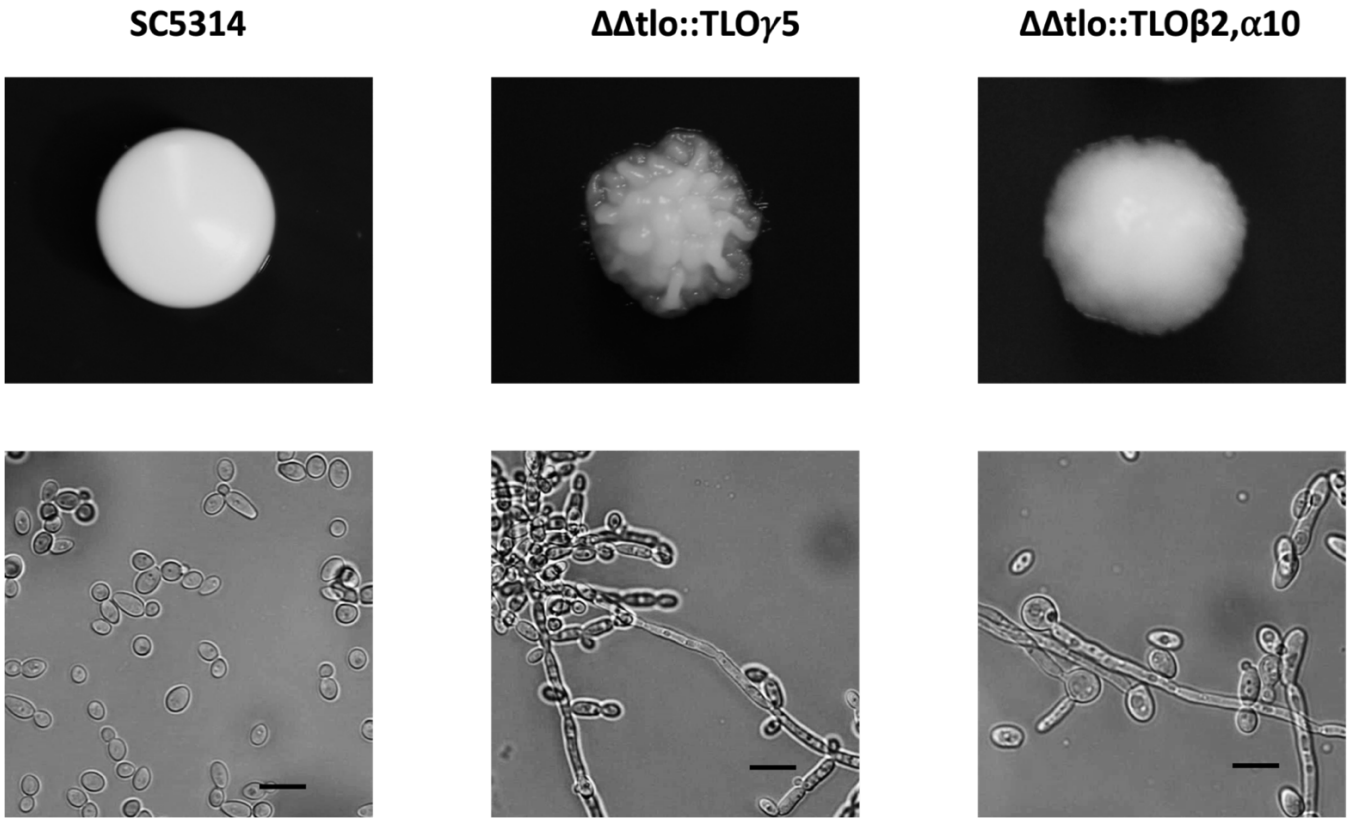
XIE, J. L. L., QIN, L. G., MIAO, Z. Q., GRYS, B. T., DIAZ, J. D., TING, K., KRIEGER, J. R., TONG, J. F., TAN, K. L., LEACH, M. D., KETELA, T., MORAN, M. F., KRYSAN, D. J., BOONE, C., ANDREWS, B. J., SELMECKI, A., WONG, K. H., ROBBINS, N. & COWEN, L. E. 2017. The *Candida albicans* transcription factor Cas5 couples stress responses, drug resistance and cell cycle regulation. *Nature Communications,* 8(1), 499.

ZAKIKHANY, K., NAGLIK, J. R., SCHMIDT-WESTHAUSEN, A., HOLLAND, G., SCHALLER, M. & HUBE, B. 2007. *In vivo* transcript profiling of *Candida albicans* identifies a gene essential for interepithelial dissemination. *Cellular Microbiology,* 9(12), 2938-2954.

ZHANG, A., PETROV, K. O., HYUN, E. R., LIU, Z., GERBER, S. A. & MYERS, L. C. 2012. The Tlo proteins are stoichiometric components of *Candida albicans* mediator anchored via the Med3 subunit. *Eukaryotic Cell,* 11(7), 874-884.

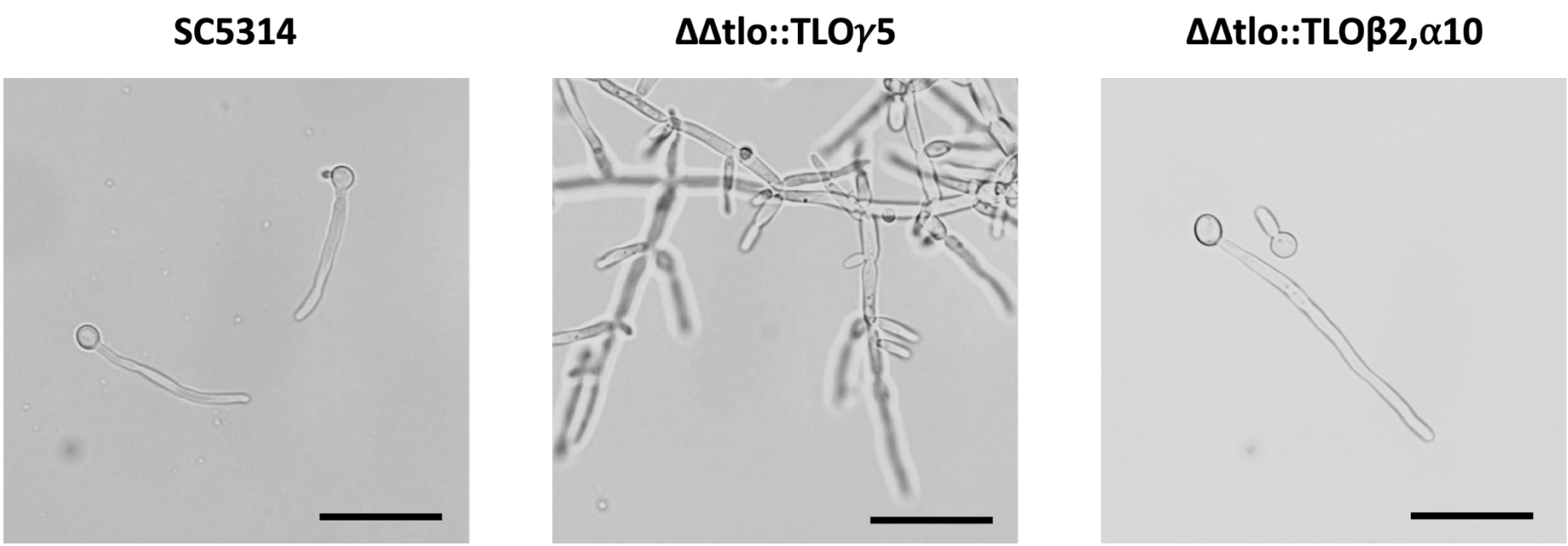
ZHANG, Y., LIU, T., MEYER, C. A., EECKHOUTE, J., JOHNSON, D. S., BERNSTEIN, B. E., NUSSBAUM, C., MYERS, R. M., BROWN, M., LI, W. & LIU, X. S. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology,* 9(9), R137.

Appendix 1 – Phenotypic analysis of *TLO* depleted strains



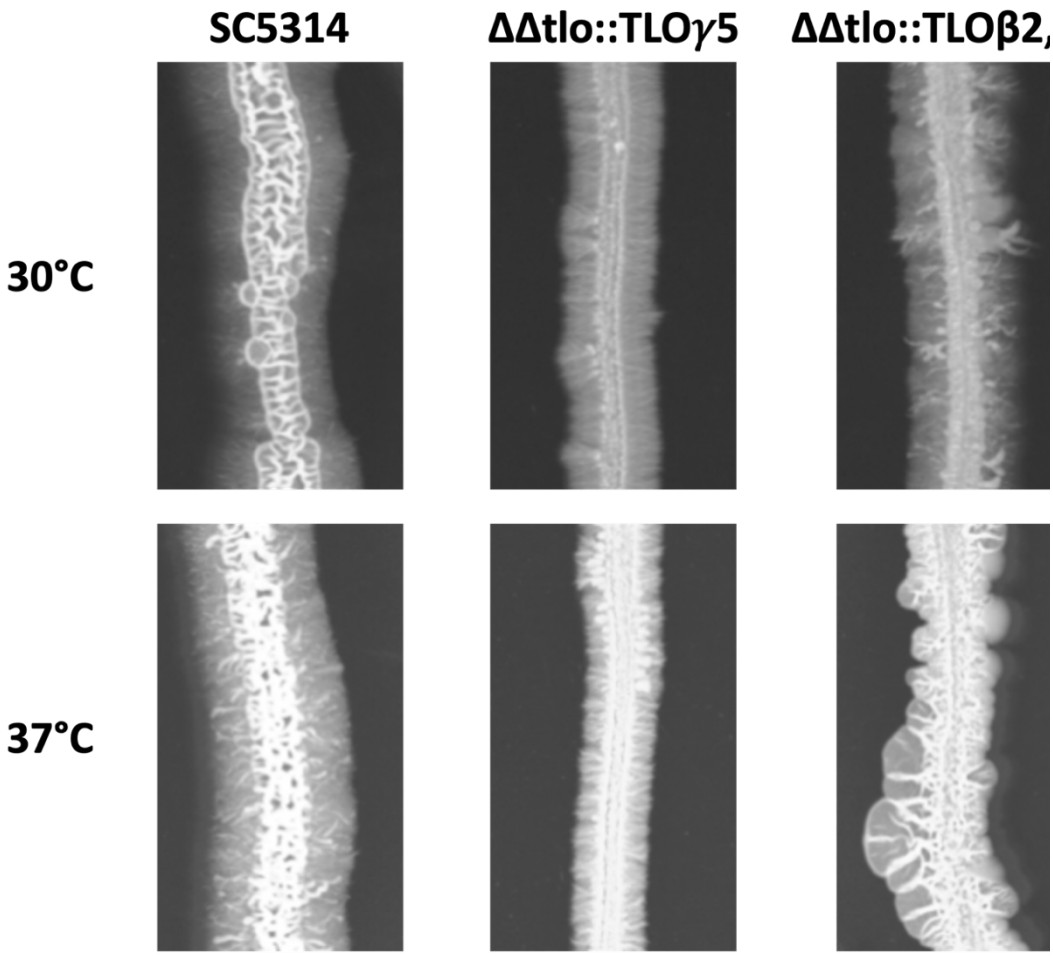
Appx. 1 Figure 1 Cellular and colony morphology of *TLO* depleted strains

Colony morphology (top) was photographed after 48h of growth on YEPD agar at 37 °C. Cellular morphology was imaged under X400 magnification after 24h of growth in YEPD liquid media. Scale-bar represents 7 μm.



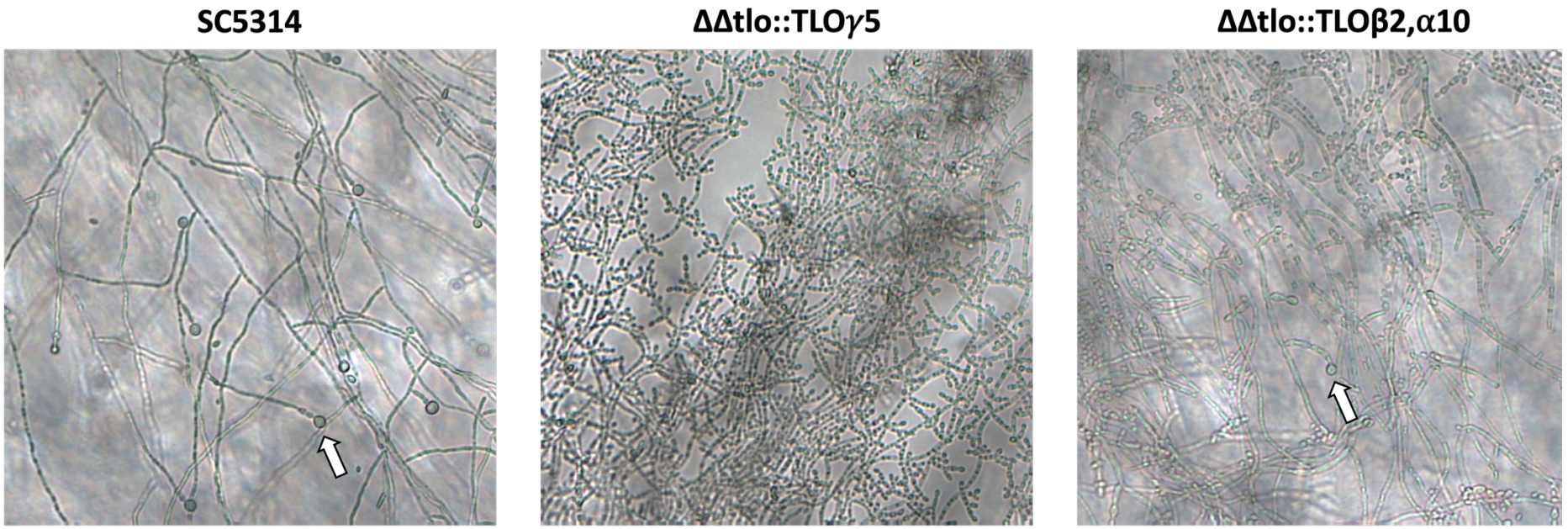
Appx. 1 Figure 2 Hyphal formation of *TLO* depleted strains in YEPD supplemented with 10% FCS

Flasks with 25 ml YEPD supplemented with 10% FCS were inoculated with 10^6 cfu/ml of each strain from cultures that had been grown overnight at 30 °C. Flasks were incubated at 37 °C in a 200 rpm shaking incubator. Cultures were monitored for hyphal formation. The images above are from the 2 h time point. Scale bar represents 7 μm.



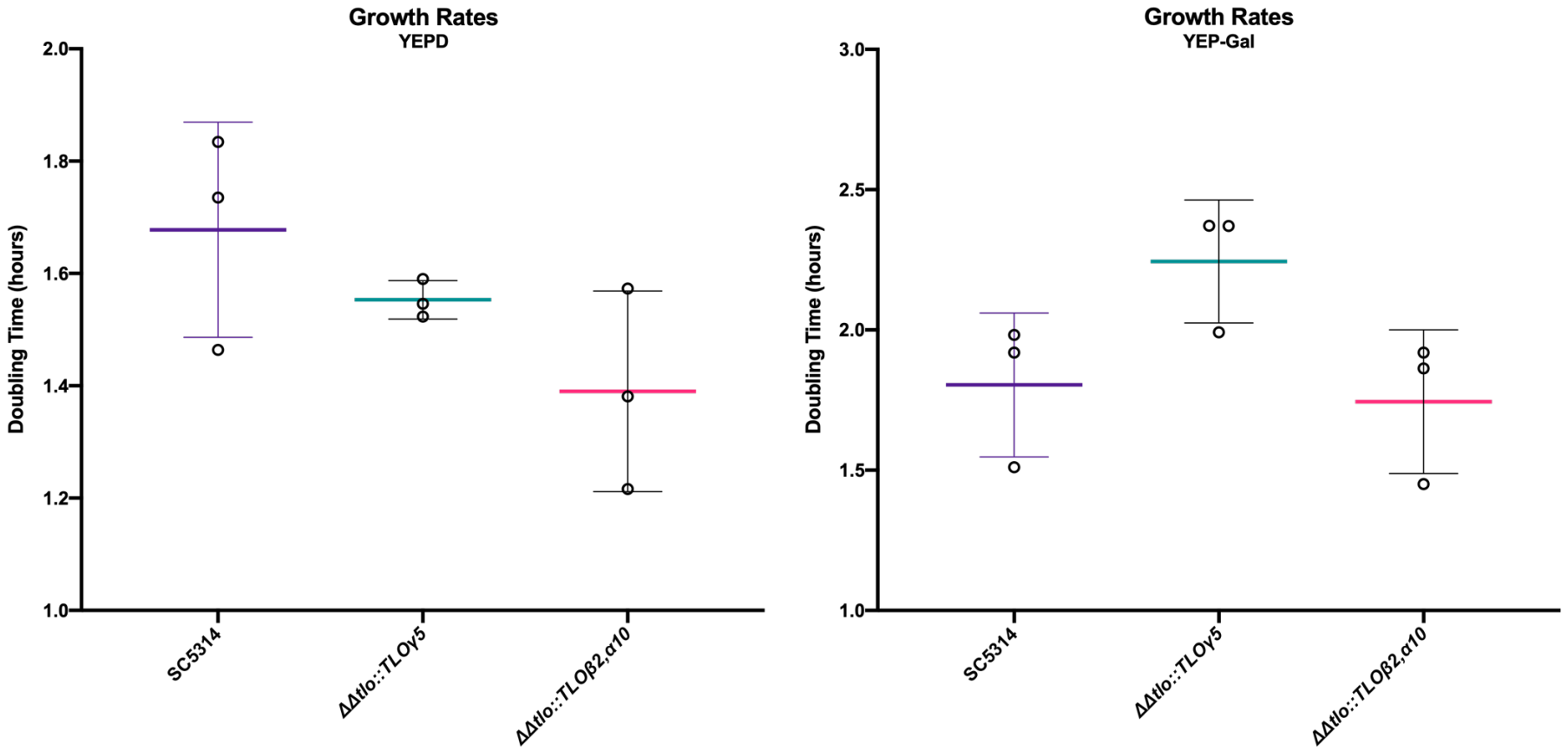
Appx. 1 Figure 3 Growth of *TLO* depleted strains on solid Spider media

The ability of strains to filament on solid hyphal media was determined by streaking a single colony of each strain onto an agar plate and incubating at either 30 °C or 37 °C for 5 days. After 5 days plates were examined and photographed.



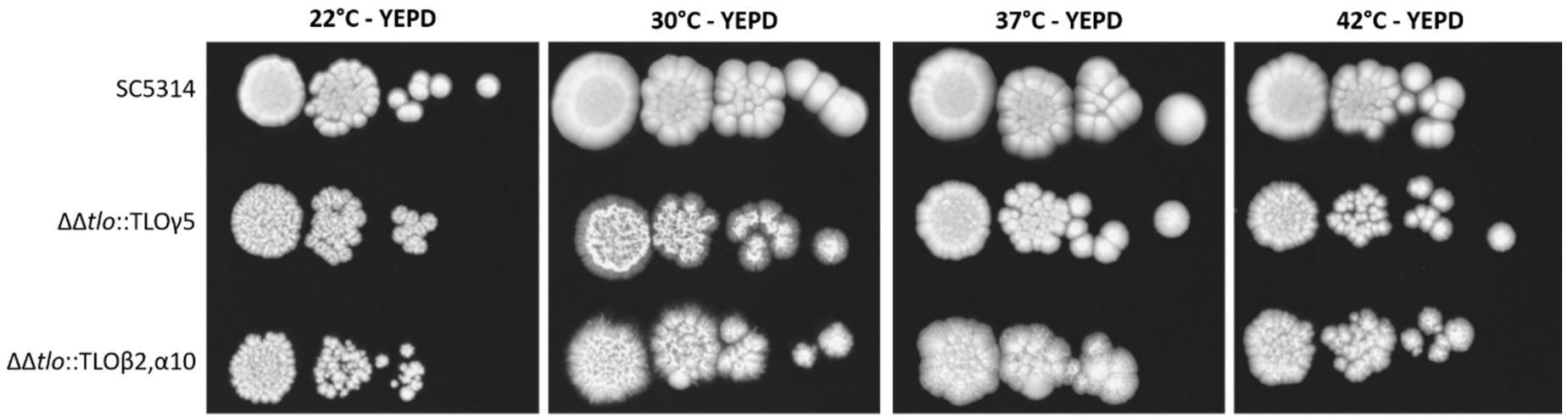
Appx. 1 Figure 4 Chlamydospore formation of *TLO* depleted strains on Corn Meal Agar with Tween 80

A single colony of each strain was streaked lightly onto corn meal agar supplemented with 1% Tween 80. Streaks were covered with a glass coverslip and the plates were incubated for 5-7 days at 22 °C in the dark. After incubation the plates were viewed and imaged directly under a light microscope. Exemplary chlamydospores are indicated by white arrows.



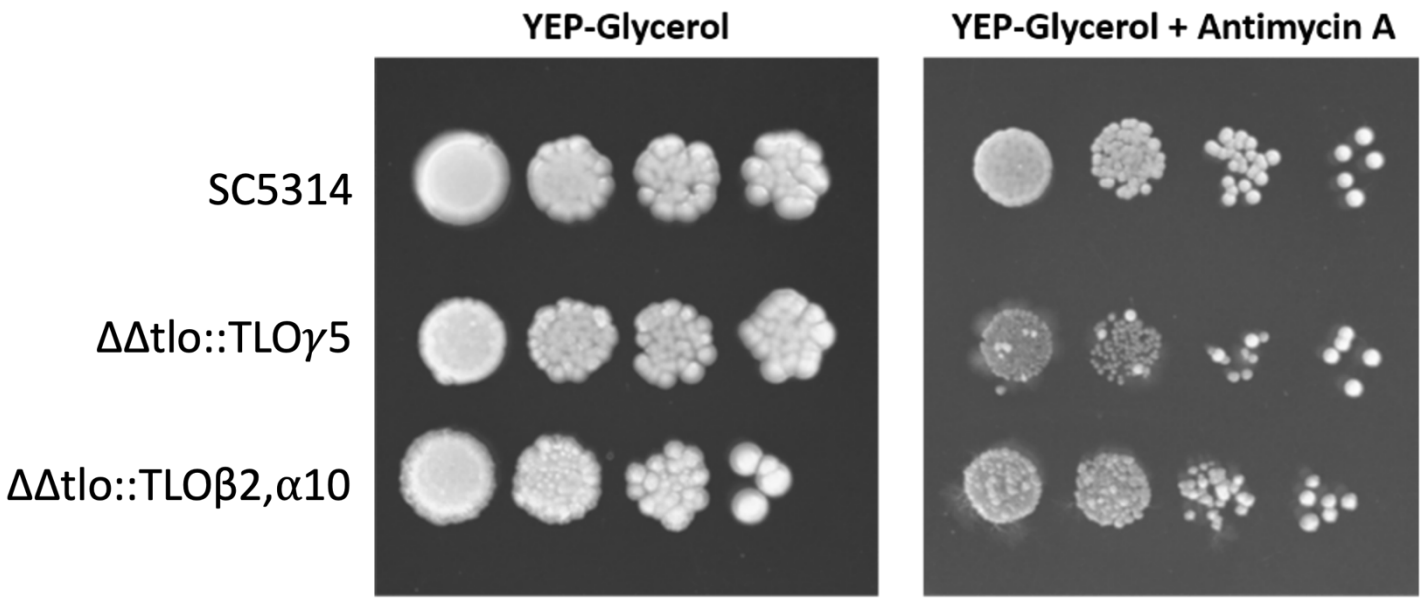
Appx. 1 Figure 5 Growth rate analysis of *TLO* depleted strains

Doubling time of each strain in wither YEPD or YEP-Galactose at 37 °C at 200 rpm was measured and is represented above by a horizontal line (hours), with error bars representing standard deviation and symbols representing each of three replicates. (A) Doubling times of strains growing in liquid YEPD. (B) Doubling times of strains growing in liquid YEP-Galactose.



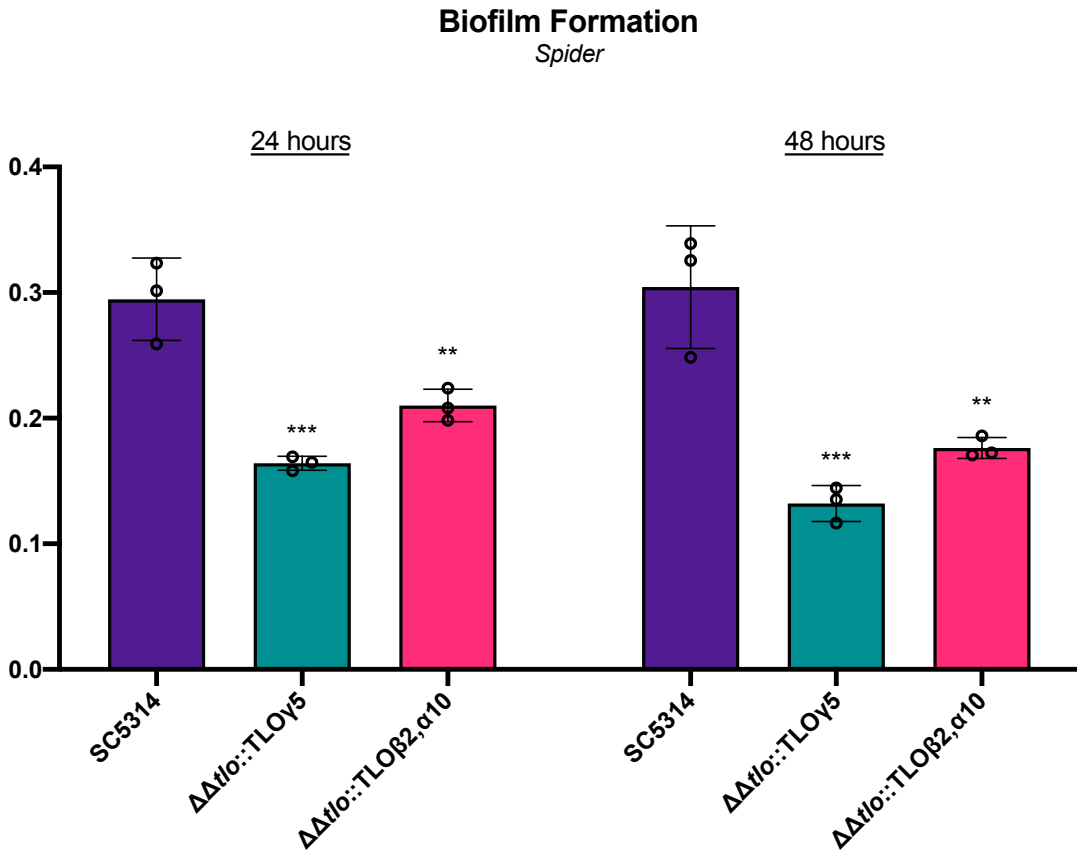
Appx. 1 Figure 6 Growth of *TLO* depleted strains on YEPD at various temperatures

10-fold serial dilutions (10^6 – 10^3 cfu/ml) were plated in 5 µl volumes onto YEPD agar and incubated for 72 h at 22, 30, 37 or 42 °C before photographing.



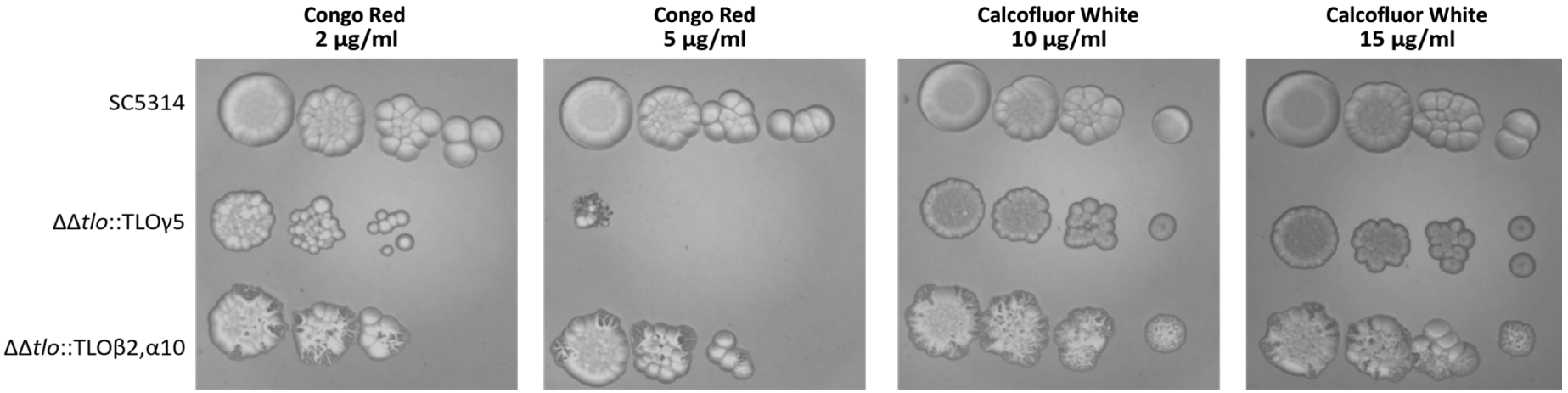
Appx. 1 Figure 7 Growth of *TLO* depleted strains in the presence of Antimycin A

10-fold serial dilutions (10^6 – 10^3 cfu/ml) were plated in 5 µl volumes onto YEP-Glycerol agar with and without 1 µg/ml Antimycin A and incubated for 72h at 37 °C before photographing.



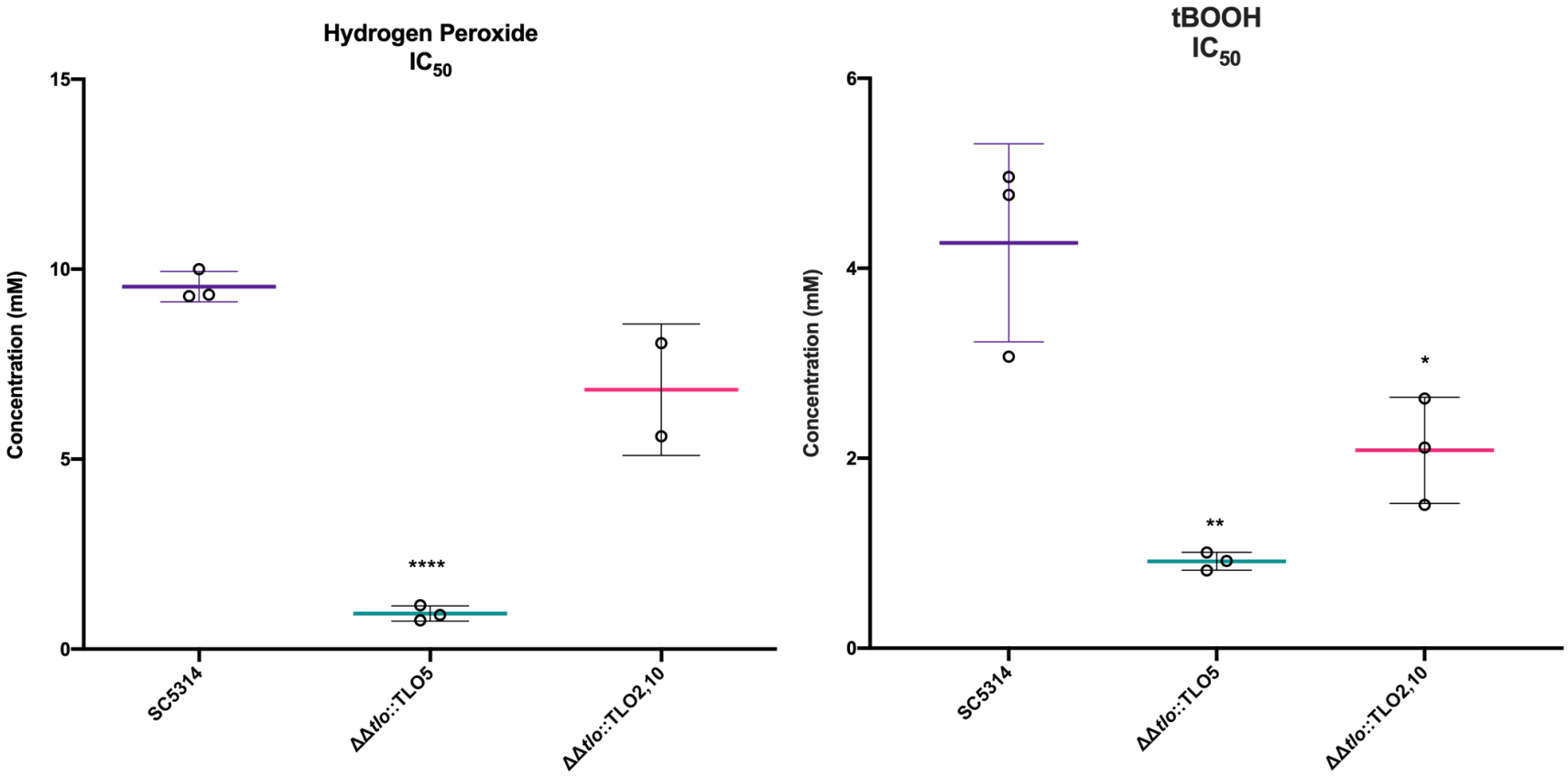
Appx. 1 Figure 8 Biofilm formation of *TLO* depleted strains

The ability of strains to form biofilm on plastic surfaces (plastic 96 well plate) in liquid Spider media was analysed. Biofilm formation was quantified by staining with crystal violet and measuring absorbance at 540 nm. This was done at 24- and 48-hour time points. (A) Graphical representation of biofilm formation data. Error bars represent standard deviation from mean, and symbols represent values from three replicates. Asterisks denote statistically significant difference from SC5314.



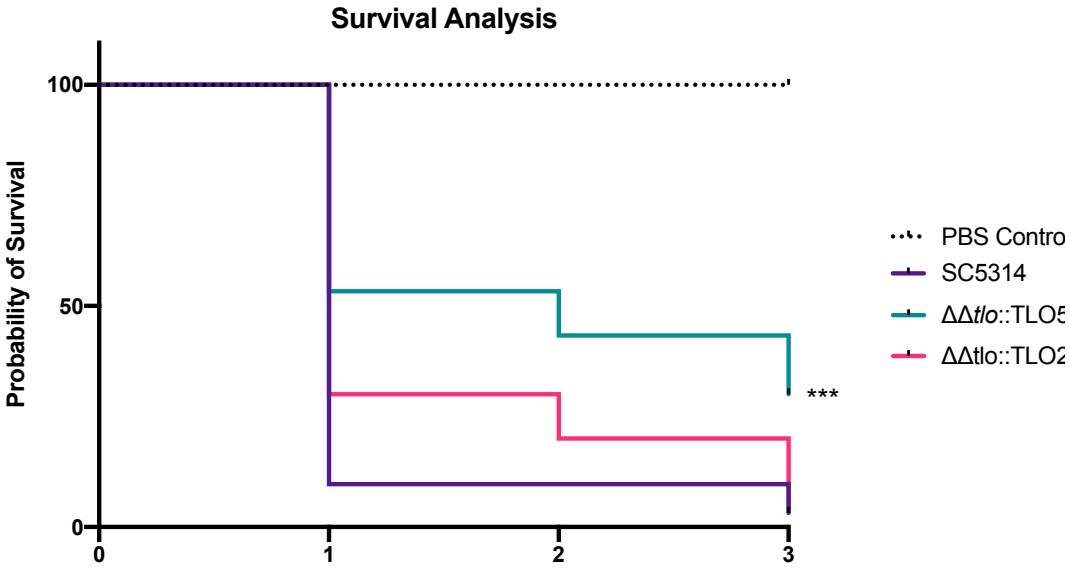
Appx. 1 Figure 9 Growth of *TLO* depleted strains on media containing cell wall perturbing compounds

10-fold serial dilutions (10^6 – 10^3 cfu/ml) were plated in 5 µl volumes onto YEPD agar containing different concentrations of either Congo Red or Calcofluor White to determine susceptibility of strains to these compounds. Plates were incubated at 37 °C for 72 h before photographing.



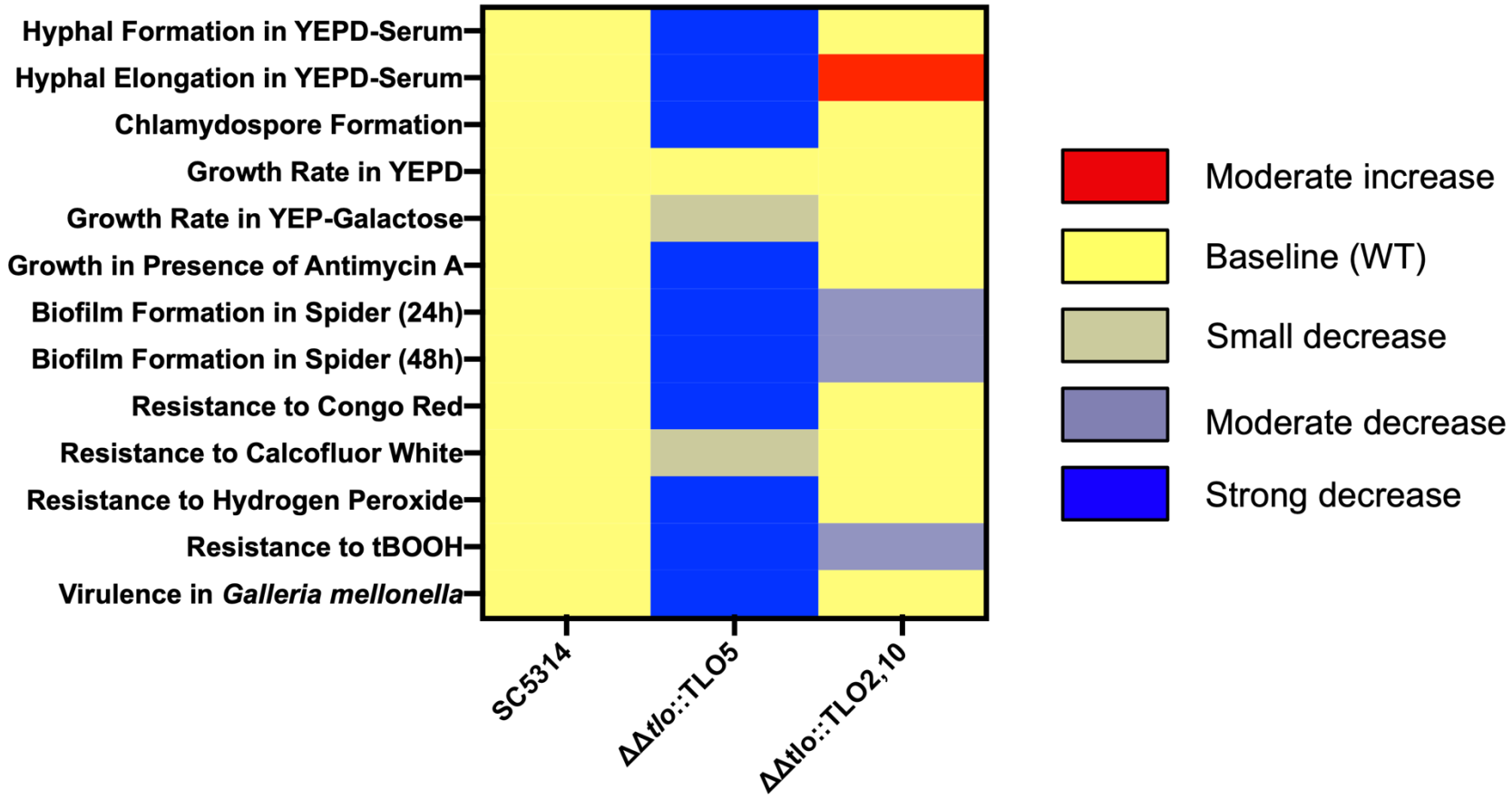
Appx. 1 Figure 10 Susceptibility of *TLO* depleted strains to oxidative stress

Minimum inhibitory concentration assays were performed by microtiter dilution to determine the IC_{50} of strains in oxidative stress inducing reagents. IC_{50} is the concentration at which growth is inhibited by 50%. (A) IC_{50} concentrations of each strain in hydrogen peroxide (H_2O_2), mean is indicated by the horizontal bar, with error bars representing standard deviation and symbols representing each of three replicates. Asterisks denote significant difference from SC5314. (B) IC_{50} concentrations of strains in tBOOH.



Appx. 1 Figure 11 Survival analysis of *Galleria mellonella* larvae after infection with *TLO* depleted strains

Each *Galleria mellonella* larva was infected with 10^6 cells in a 20µl volume. Inoculation with the same volume of PBS was used as a negative control. In total 20 worms were infected with each strain/PBS and a Kaplan-Meyer curve of survival of larvae at 24 hour time points over 72 h was generated.



Appx. 1 Figure 12. Heatmap of phenotypes observed in *TLO* depletion strains

Phenotypes were graded compared to the response of the SC5314 WT strain. Grading scale can be seen on the right hand side.

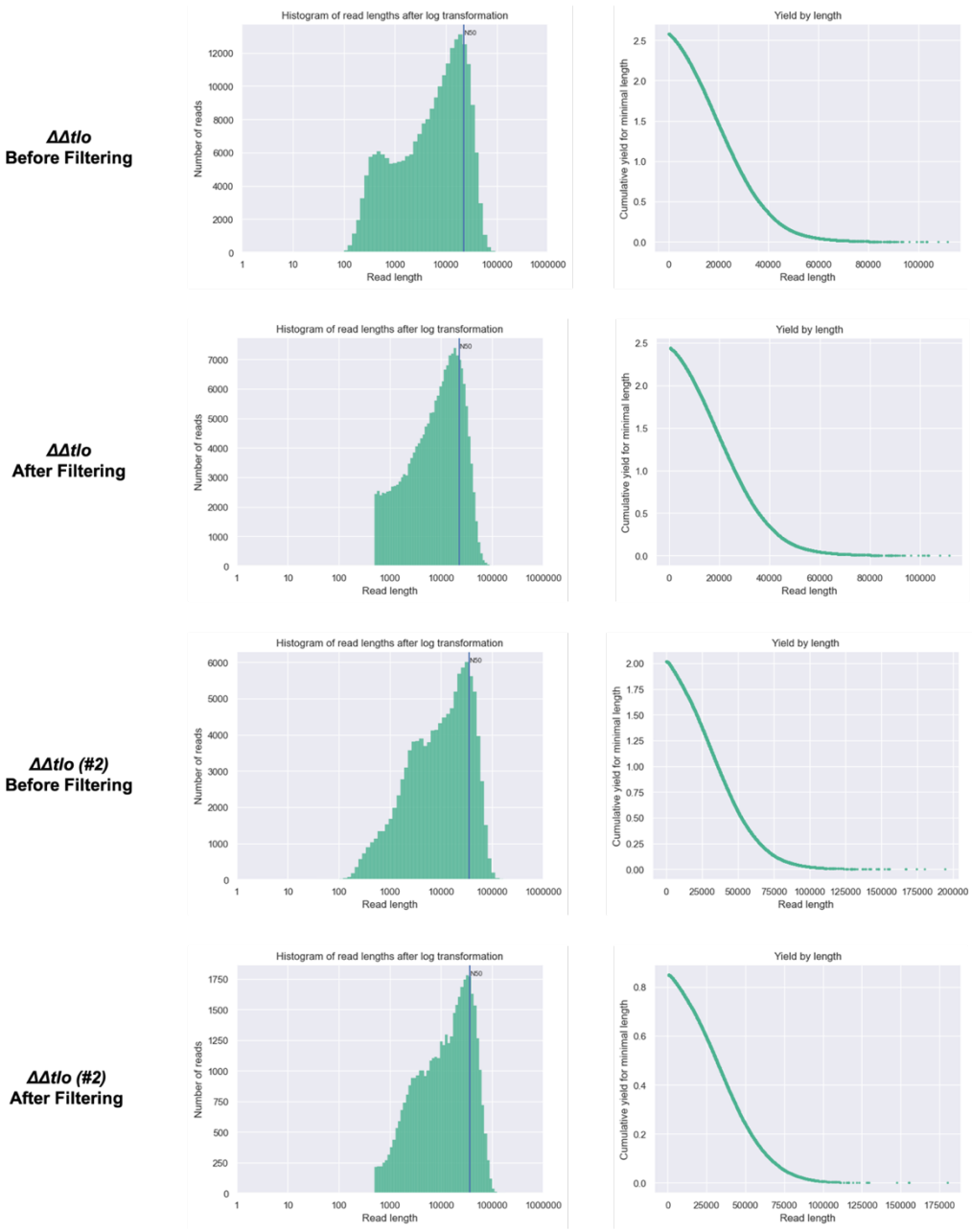
Appendix 2 – Quality Statistics from WGS of *ΔΔtlo* strains

	ΔΔtlo Before Filtering	*ΔΔtlo* After Filtering	*ΔΔtlo* (#2) Before Filtering	*ΔΔtlo* (#2) After Filtering
Mean read length	10,970.70	12,830.10	17,926.60	19,947.80
Mean read quality	9.3	9.7	9.3	10.6
Median read length	6,456.00	8,991.00	10,603.00	13,325.50
Median read quality	9.6	9.8	9.7	10.5
Number of reads	234,778.00	190,079	112,457.00	42,582.00
Read length N50	22,491.00	22,694.00	35,553.00	36,321.00
Total bases	2,575,688,384.00	2,438,740,044.00	2,015,971,621.00	849,418,085.00
Number, percentage and megabases of reads above quality cut-offs				
>Q5	233212 (99.3%) 2567.5 Mb	190079 (100%) 2438.7 Mb	111316 (99.0%) 2004.8Mb	42582 (100%) 849.4 Mb
>Q7	218800 (93.2%) 2468.3 Mb	189985 (100%) 2438.7 Mb	103443 (92.0%) 1903.8 Mb	42582 (100%) 849.4 Mb
>Q10	87719 (37.4%) 1125.4 Mb	81060 (42.6%) 1120.9 Mb	4406 (39.1%) 854.2 Mb	42324 (99.4%) 848.0 Mb
>Q12	349 (0.1%) 0.4 Mb	126 (0.1%) 0.4 Mb	83 (0.1%) 0.1 Mb	48 (0.1%) 0.1 Mb
Average Coverage	175.2	76.5	137.2	57.8

Appx. 2 Table 1 Output of NanoStat QC analysis for fastq files before and after filtering cont.

	ΔΔtlo Before Filtering	*ΔΔtlo* After Filtering	*ΔΔtlo* (#2) Before Filtering	*ΔΔtlo* (#2) After Filtering
Top 5 highest mean basecall quality scores and their read lengths				
1	14.3 (152)	12.7 (536)	13.7 (177)	12.9 (855)
2	14.2 (139)	12.6 (640)	13.4 (275)	12.6 (509)
3	14.0 (216)	12.6 (799)	13.0 (585)	12.6 (617)
4	13.8 (397)	12.6 (645)	12.9 (188)	12.5 (1909)
5	13.7 (146)	12.6 (990)	12.8 (399)	12.5 (553)
Top 5 longest reads and their mean basecall quality score				
1	111518 (10.2)	111518 (10.2)	194334 (9.7)	179933 (10.4)
2	107982 (10.6)	107982 (10.6)	180033 (10.4)	155433 (10.5)
3	103593 (9.6)	103593 (9.6)	174489 (7.6)	154974 (10.3)
4	103219 (8.7)	103219 (8.7)	167187 (8.0)	147667 (10.1)
5	103102 (8.0)	103102 (8.0)	166784 (9.7)	129459 (10.2)

Appx. 2 Table 1 Output of NanoStat QC analysis for fastq files before and after filtering cont.

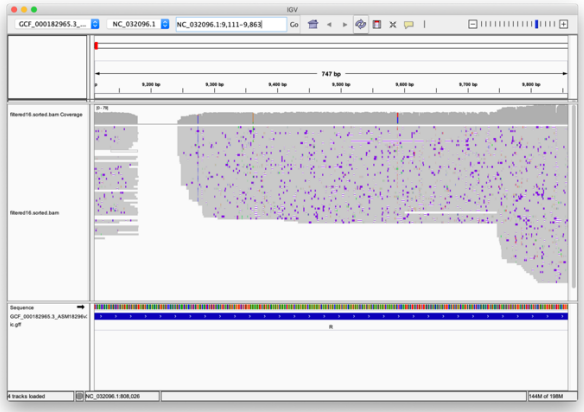


Appx. 2 Figure 1 NanoQC output graphs for *ΔΔtlo* whole genome sequencing

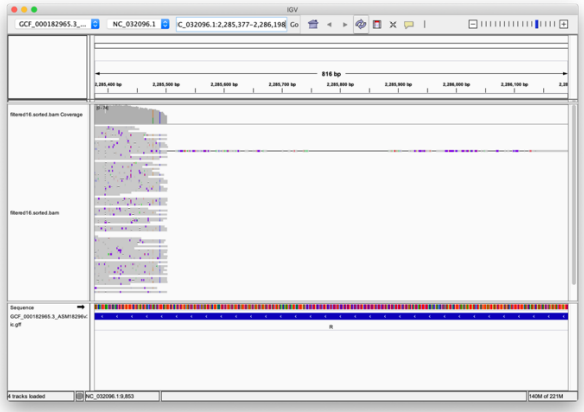
Representitive graphs from the NanoQC tool detailing the length of reads in the fastq files for the *ΔΔtlo* and *ΔΔtlo* (#2) strains before and after filtering based on length (min=50) and quality (min=Q10).

Appendix 3 – WGS results for the *ΔΔtlo* strains

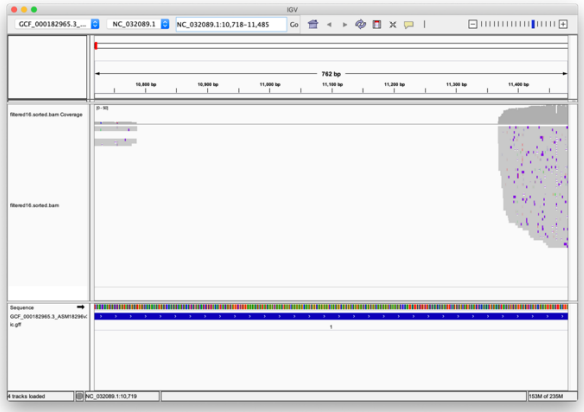
TLOα1



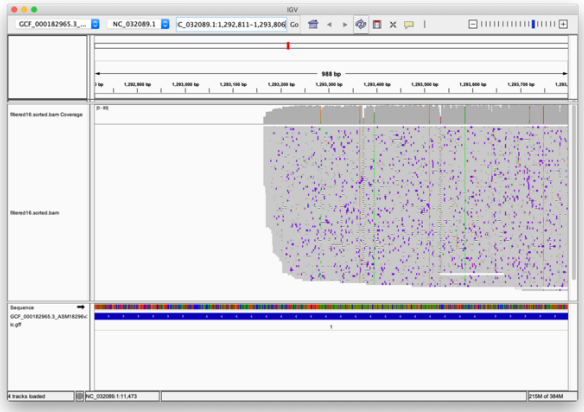
TLOβ2



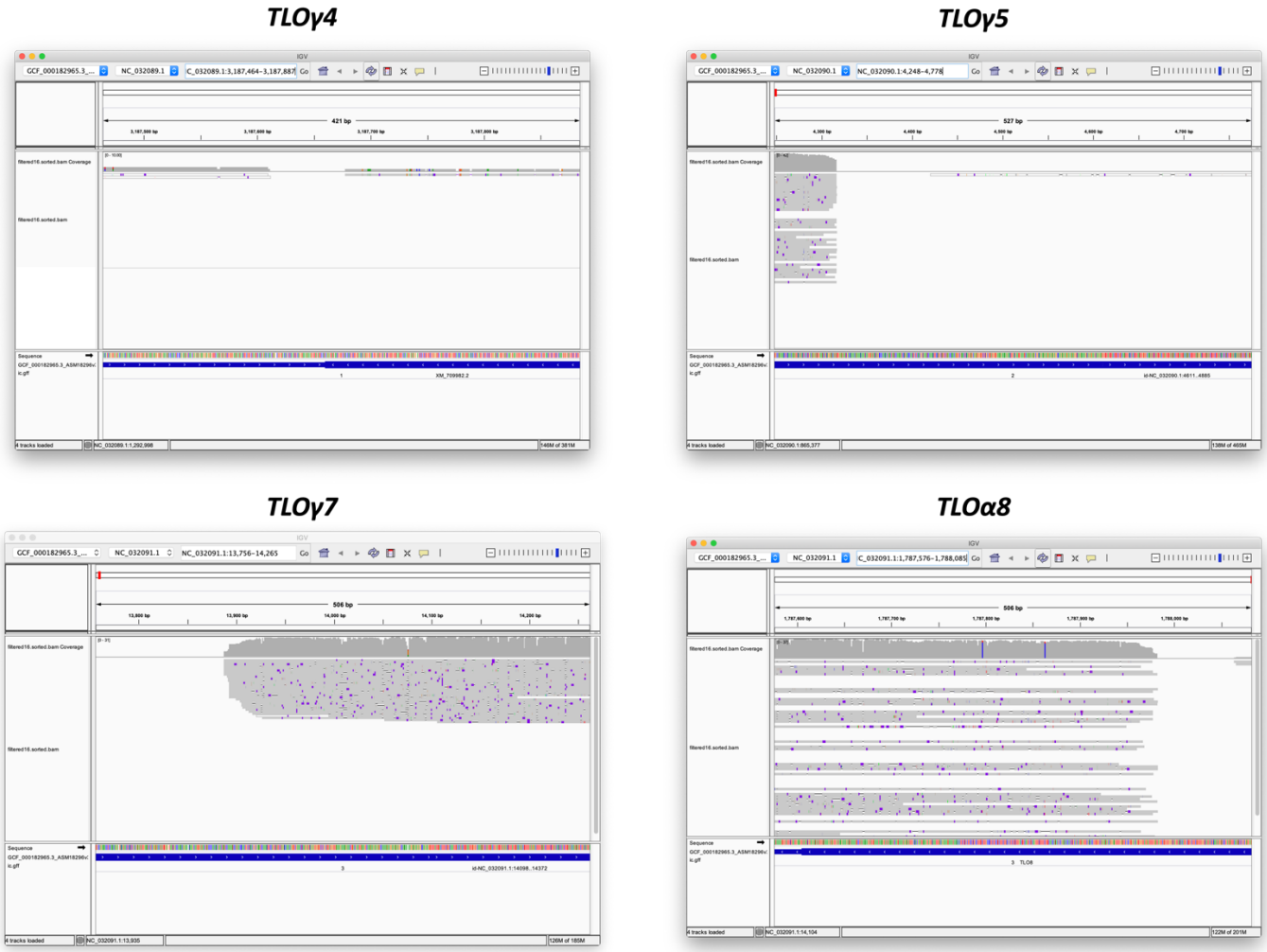
TLOα3



TLOα34

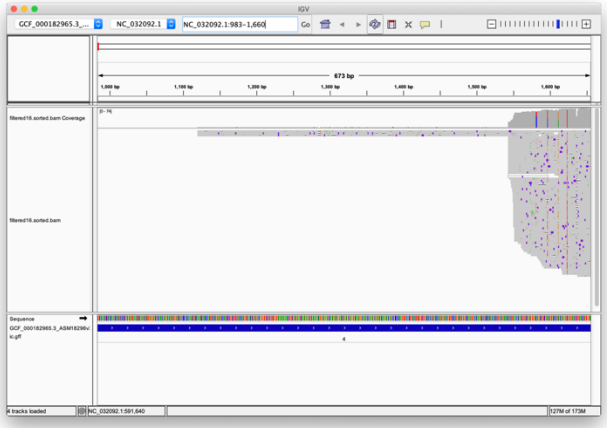


Appx. 3 Figure 1 *TLO* deletions visualised in IGV for *ΔΔtlo* strain (cont.)

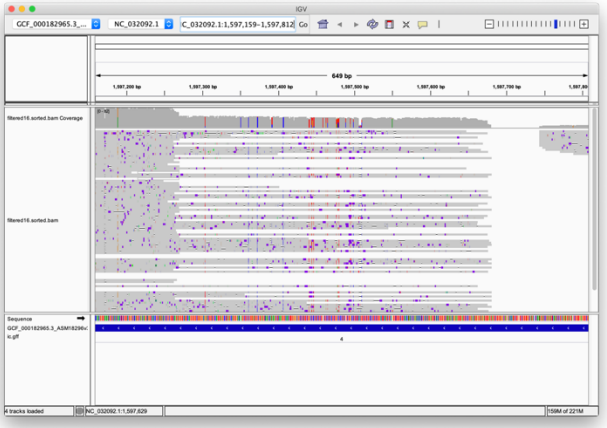


Appx. 3 Figure 1 *TLO* deletions visualised in IGV for *ΔΔtlo* strain (cont.)

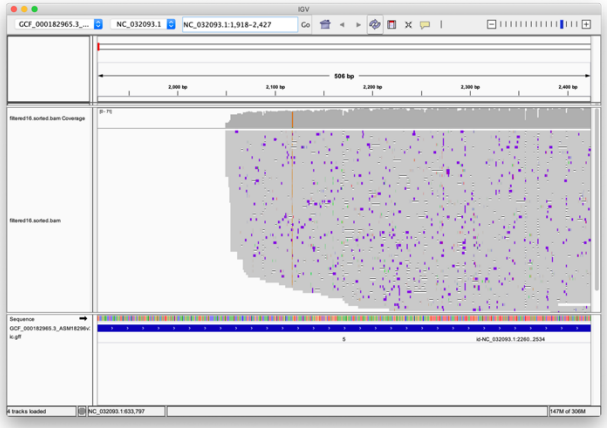
TLOα9



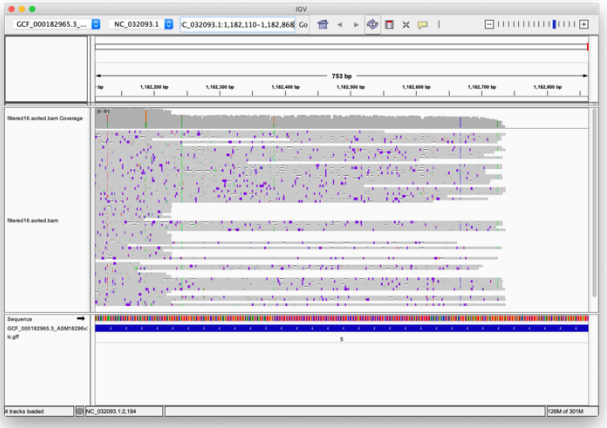
TLOα10



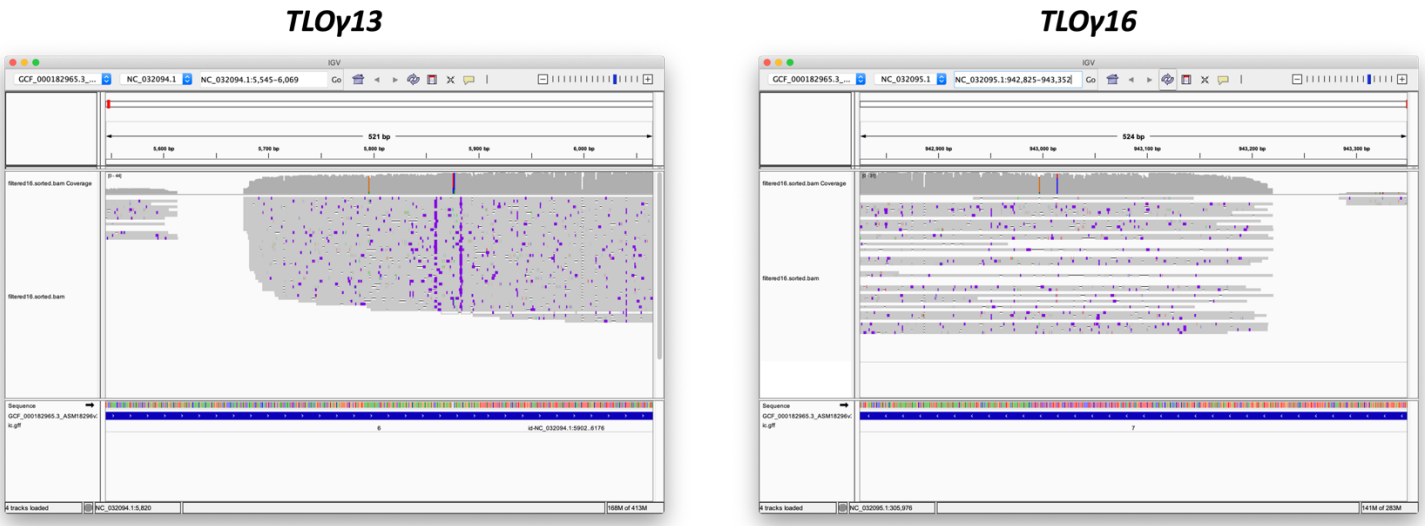
TLOγ11



TLOα12

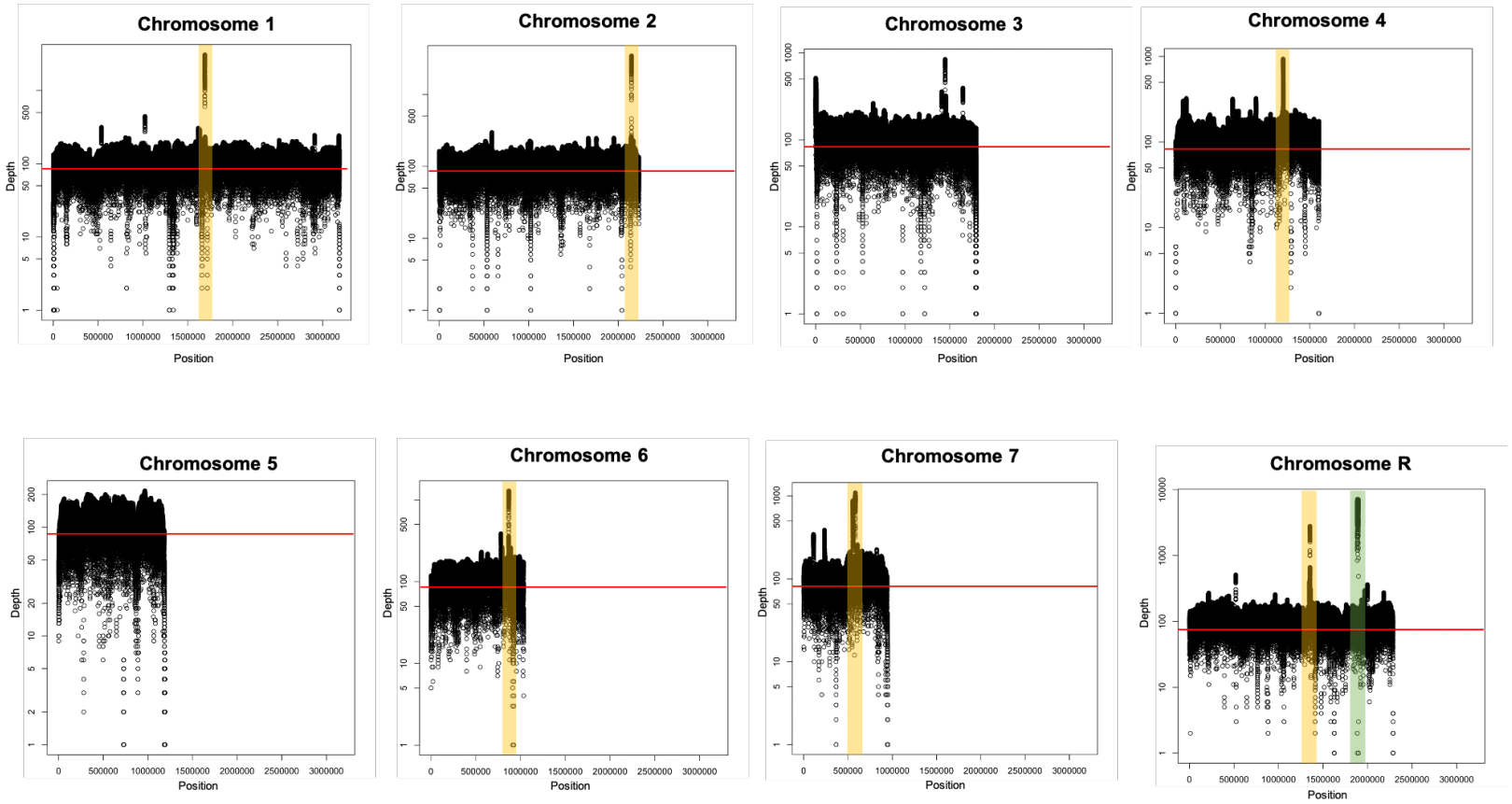


Appx. 3 Figure 1 *TLO* deletions visualised in IGV for *ΔΔtlo* strain (cont.)



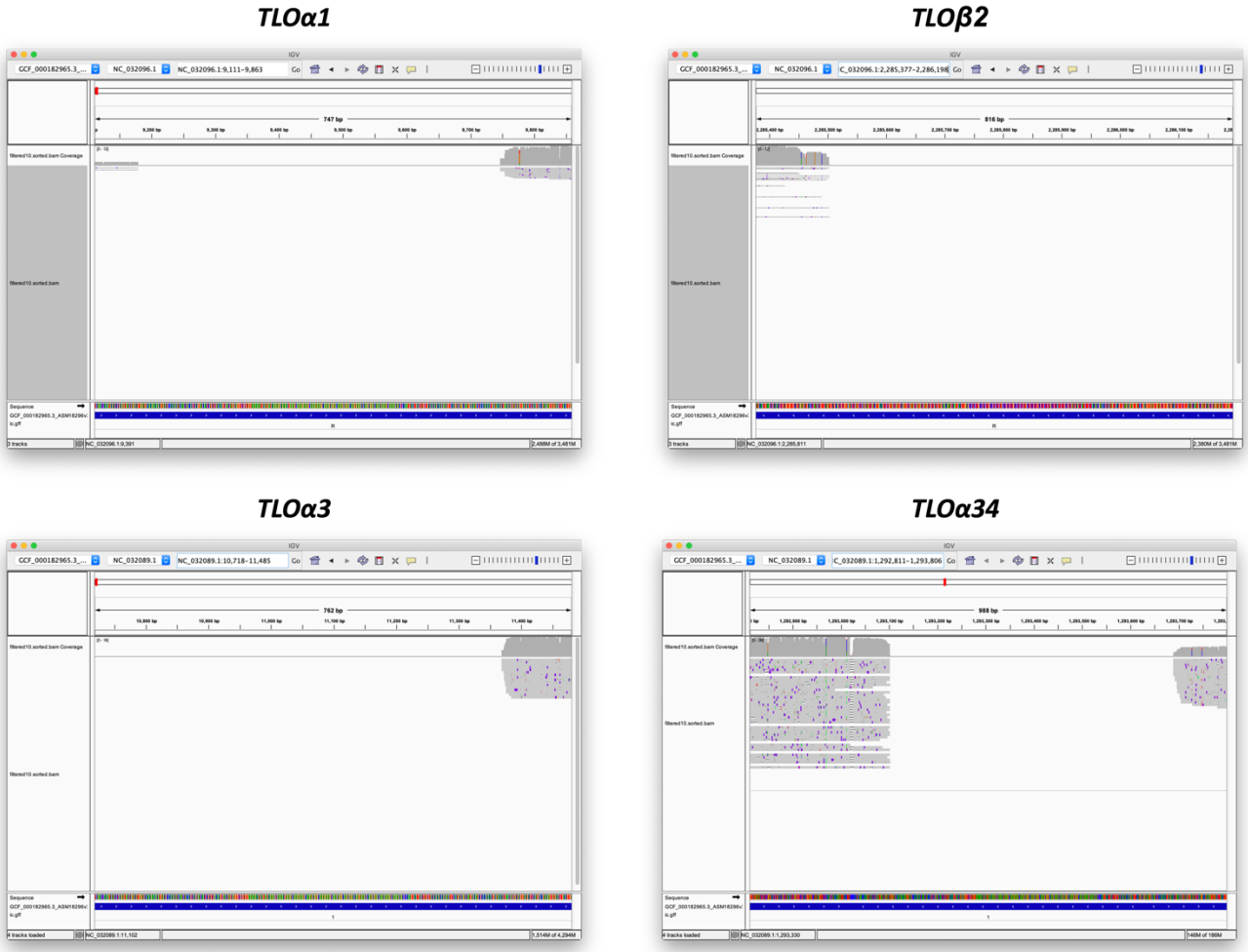
Appx. 3 Figure 1 *TLO* deletions visualised in IGV for *ΔΔtlo* strain (cont.)

The primary *TLO* null mutant strain (*ΔΔtlo*), with the cassette recycled, was sequenced using Oxford Nanopore MinION technology and aligned to the reference genome. This alignment was visualised using the Integrated Genome Viewer (IGV). The coordinates for each *TLO* gene were inputted and a screenshot of the read pileup was taken. Each window represents the gene body of each *TLO* gene. The top track labelled "filtered16.sorted.bam.Coverage" displays a histogram of the coverage at each base along the displayed segment. The track labelled "filtered16.sorted.bam" shows the aligned reads as grey bars, with purple highlights indicating deviations from the reference sequence. The bottom track is the GFF reference file, which shows the locations of genes and ORFs, which are indicated by the blue bars, and the directionality is indicated by white arrows in the blue bars.

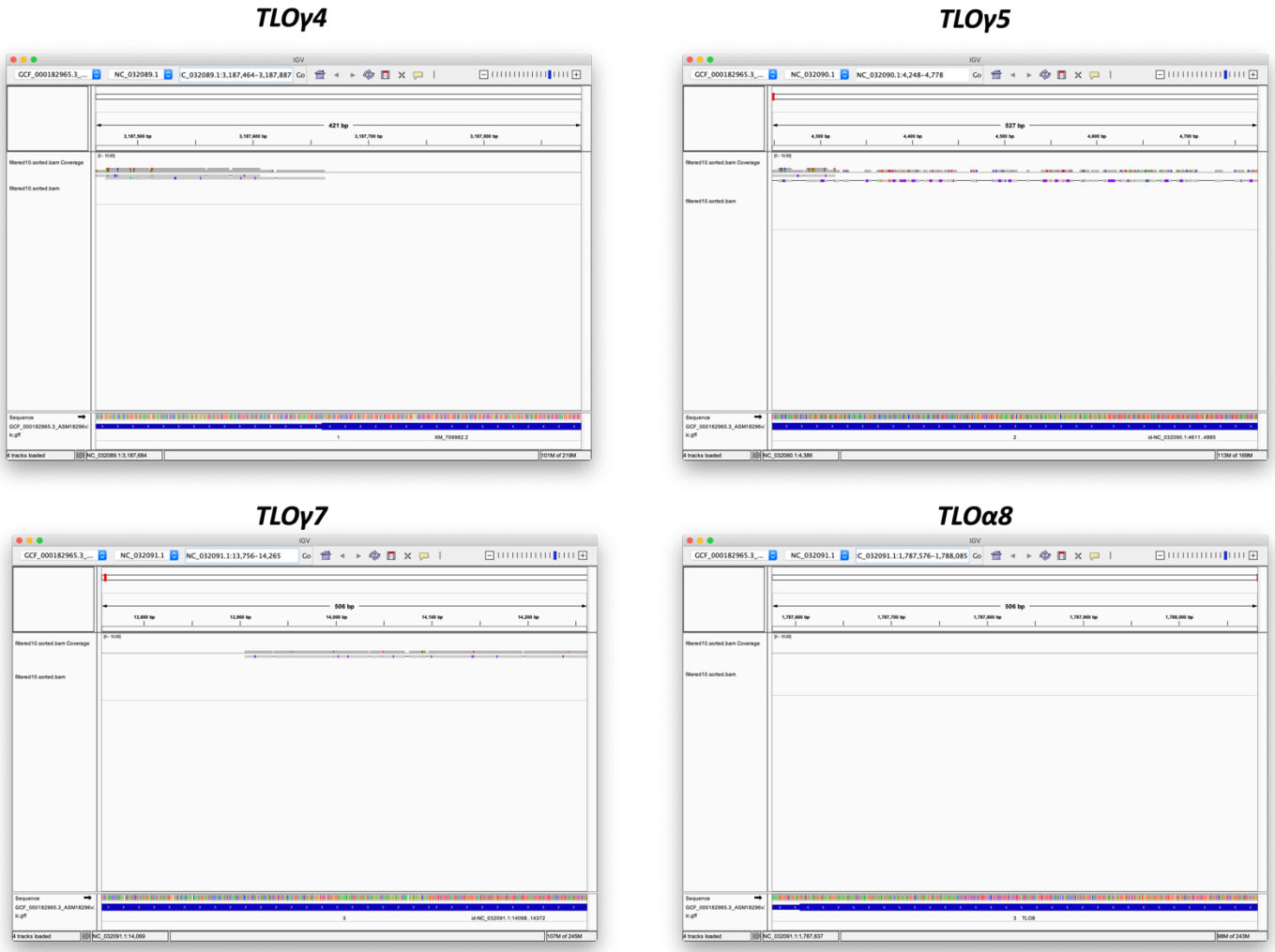


Appx 3. Figure 2 Whole genome sequencing coverage plots for the *ΔΔtlo* strain

The coverage at each base was generated from the aligned *ΔΔtlo* file using SAMtools and plotted using R. The red bar across the plot represents the theoretical coverage across the whole genome (generated from statistical analysis of the fastq file). The yellow boxes indicate the Major Repeat Sequences (MRS). The highly repetitive nature of these regions can result in incorrect pileup when aligning the reads to the reference. The green box denotes the rDNA locus on chromosome R, another highly repetitive region.

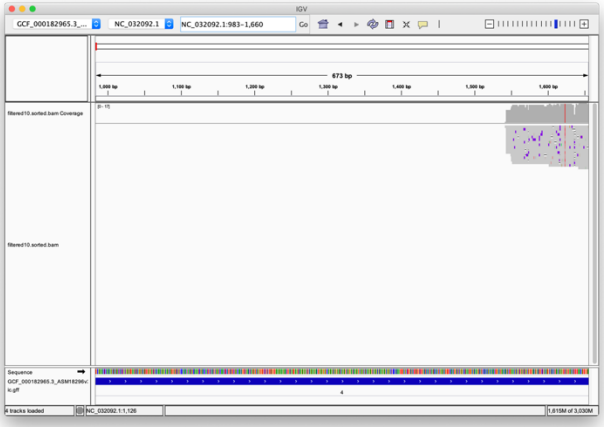


Appx. 3 Figure 3 *TLO* deletions visualised on IGV for *ΔΔtlo* strain (#2) (cont.)

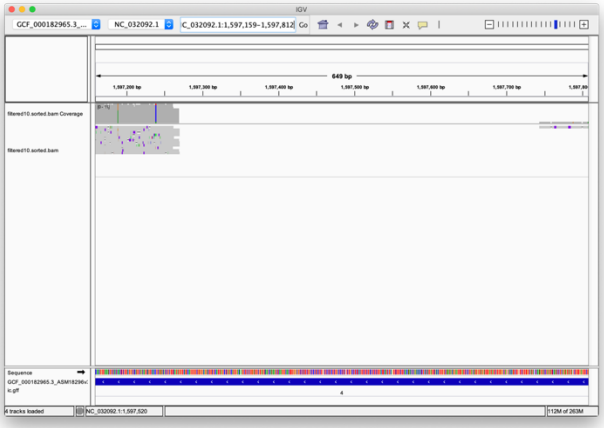


Appx. 3 Figure 3 *TLO* deletions visualised on IGV for *ΔΔtlo* strain (#2) (cont.)

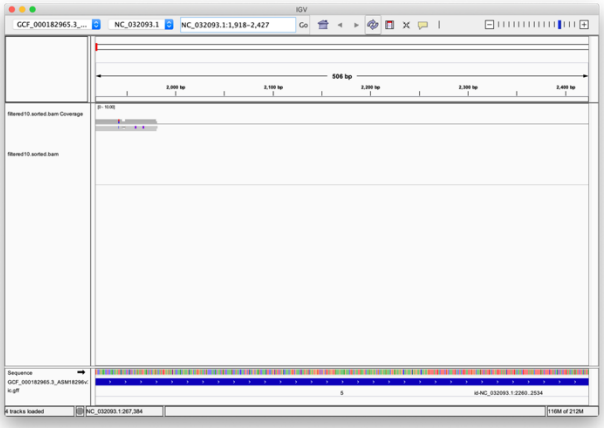
TLOα9



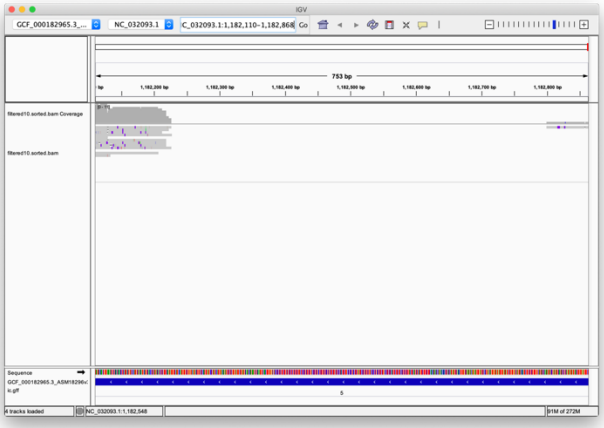
TLOα10



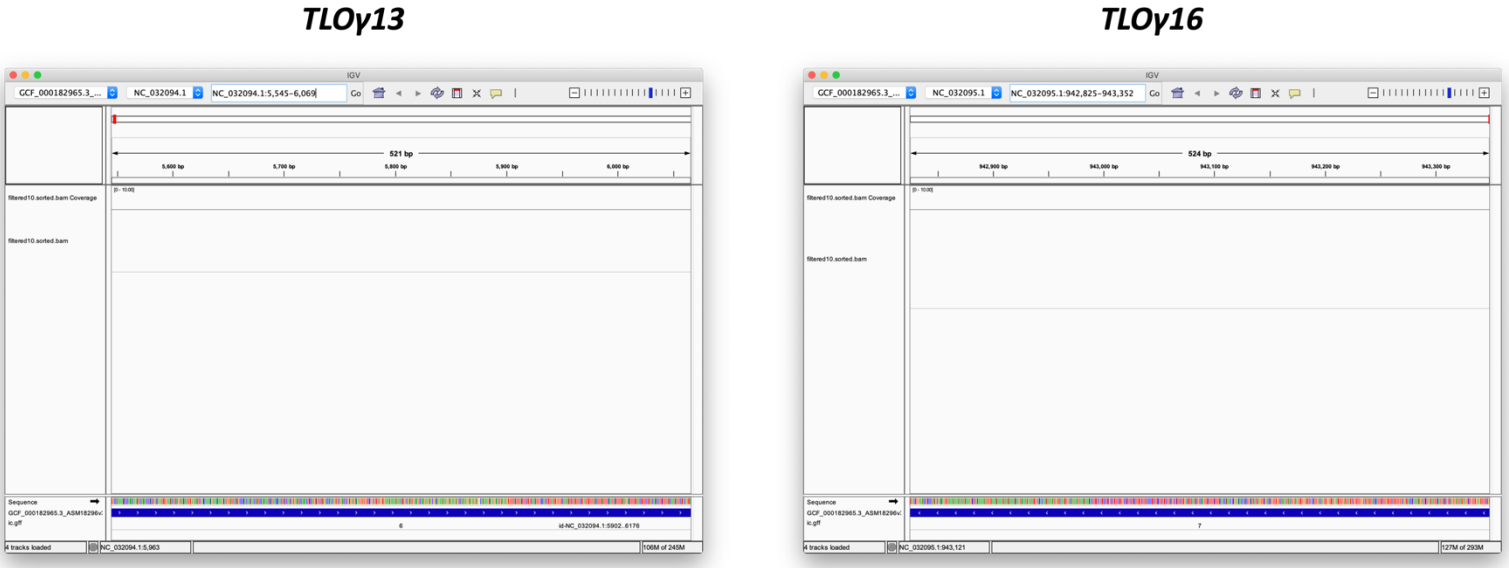
TLOγ11



TLOα12

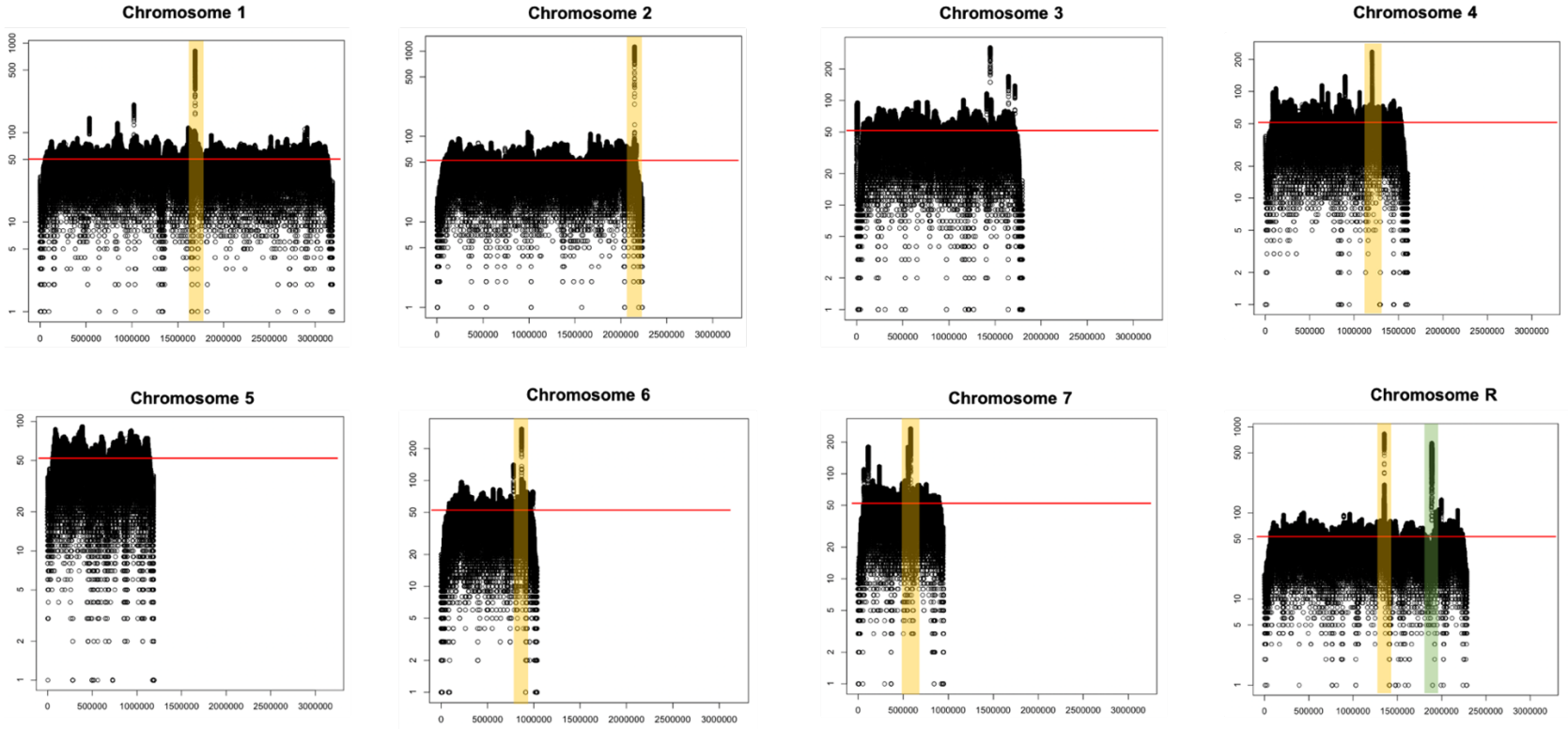


Appx. 3 Figure 3 *TLO* deletions visualised on IGV for *ΔΔtlo* strain (#2) (cont.)



Appx. 3 Figure 3 *TLO* deletions visualised on IGV for *ΔΔtlo* strain (#2) (cont.)

The second *TLO* null mutant strain (*ΔΔtlo* #2), with the cassette recycled, was sequenced using Oxford Nanopore MinION technology and aligned to the reference genome. This alignment was visualised using the Integrated Genome Viewer (IGV). The coordinates for each *TLO* gene were inputted and a screenshot of the read pileup was taken. Each window represents the gene body of each *TLO* gene. The top track labelled "filtered10.sorted.bam.Coverage" displays a histogram of the coverage at each base along the displayed segment. The track labelled "filtered10.sorted.bam" shows the aligned reads as grey bars, with purple highlights indicating deviations from the reference sequence. The bottom track is the GFF reference file, which shows the locations of genes and ORFs, which are indicated by the blue bars, and the directionality is indicated by white arrows in the blue bars.



Appx. 3 Figure 4 Whole genome sequencing coverage plots for the *ΔΔtlo* (#2) strain

The coverage at each base was generated from the aligned *ΔΔtlo* (#2) file using SAMtools and plotted using R. The red bar across the plot represents the theoretical coverage across the whole genome (generated from statistical analysis of the fastq file). The yellow boxes indicate the Major Repeat Sequences (MRS). The highly repetitive nature of these regions can result in incorrect pileup when aligning the reads to the reference. The green box denotes the rDNA locus on chromosome R, another highly repetitive region.

Appendix 4 – Quality statistics for RNA sequencing

Sample name	Raw reads	Raw bases	Error rate(%)	Q30(%)	GC content(%)
AHY940 0h (1)	20355491	6.1	0.03	90.68	36.75
AHY940 0h (2)	22541629	6.8	0.02	94.18	36.78
AHY940 0h (3)	21975238	6.6	0.03	93.94	37.18
AHY940 30m (1)	59556701	17.9	0.03	93.58	36.52
AHY940 30m (2)	21828217	6.5	0.02	94.24	36.58
AHY940 30m (3)	20969592	6.3	0.02	94.19	36.77
ΔΔtlo 0h (1)	22279378	6.7	0.03	92.63	36.67
ΔΔtlo 0h (2)	21718407	6.5	0.02	94.05	36.91
ΔΔtlo 0h (3)	22658284	6.8	0.03	94.04	36.86
ΔΔtlo 30m (1)	20409707	6.1	0.03	93.94	36.55
ΔΔtlo 30m (2)	22533940	6.8	0.03	94.02	36.76
ΔΔtlo 30m (3)	22017225	6.6	0.03	93.94	36.56
TLOα1 0h (1)	22130265	6.6	0.03	94.07	37
TLOα1 0h (2)	22688616	6.8	0.03	93.5	37.1
TLOα1 0h (3)	24160349	7.2	0.03	93.85	37.26
TLOα1 30m (1)	21456985	6.4	0.03	93.91	36.45
TLOα1 30m (2)	22860798	6.9	0.03	91.23	36.39
TLOα1 30m (3)	20918819	6.3	0.02	94.14	36.26
TLOβ2 0h (1)	21101487	6.3	0.03	93.48	36.82
TLOβ2 0h (2)	20741110	6.2	0.03	93.93	36.82
TLOβ2 0h (3)	21484831	6.4	0.02	94.13	36.67
TLOβ2 30m (1)	21737701	6.5	0.02	94.5	36.59
TLOβ2 30m (2)	22230214	6.7	0.03	93.93	36.2
TLOβ2 30m (3)	23397038	7	0.03	93.8	36.59
TLOγ11 0h (1)	22179911	6.7	0.03	94	36.7
TLOγ11 0h (2)	21999187	6.6	0.03	93.6	36.78
TLOγ11 0h (3)	20118037	6	0.03	93.25	36.58
TLOγ11 30m (1)	23217768	7	0.03	92.66	36.74
TLOγ11 30m (2)	21355976	6.4	0.03	93.33	36.64
TLOγ11 30m (3)	19835329	6	0.03	90.51	36.68

Appx. 4 Table 1 Quality statistics of RNA-sequencing samples

The 0 hour time points indicate that the strain was grown to mid-exponential phase in YEPD, the 30m time points were harvested after 30 minutes of exposure to tBOOH.

Sample name	Total reads	Total mapped reads	Total mapping rate	Uniquely mapping rate	Multiple mapping rate
AHY940 0h (1)	40148898	38196361	95.14%	93.01%	2.13%
AHY940 0h (2)	44705696	43318923	96.90%	94.96%	1.94%
AHY940 0h (3)	43328940	42009027	96.95%	95.02%	1.93%
AHY940 30m (1)	14319284	13859217	96.79%	94.37%	2.41%
AHY940 30m (2)	43263648	41948250	96.96%	94.87%	2.09%
AHY940 30m (3)	40836088	38908637	95.28%	91.80%	3.48%
ΔΔtlo 0h (1)	44012172	42467387	96.49%	94.34%	2.15%
ΔΔtlo 0h (2)	42262990	40889777	96.75%	94.53%	2.22%
ΔΔtlo 0h (3)	44709470	43434756	97.15%	94.94%	2.21%
ΔΔtlo 30m (1)	40215084	38706836	96.25%	93.06%	3.19%
ΔΔtlo 30m (2)	43897420	42199148	96.13%	92.69%	3.45%
ΔΔtlo 30m (3)	43588382	42091619	96.57%	93.69%	2.87%
TLOα1 0h (1)	43929382	42739566	97.29%	95.04%	2.25%
TLOα1 0h (2)	44904646	43618541	97.14%	94.99%	2.15%
TLOα1 0h (3)	47925180	46726670	97.50%	95.11%	2.39%
TLOα1 30m (1)	41891438	40339962	96.30%	92.93%	3.36%
TLOα1 30m (2)	45161422	42528524	94.17%	90.46%	3.71%
TLOα1 30m (3)	41220176	39359414	95.49%	91.50%	3.99%
TLOβ2 0h (1)	41706362	40515280	97.14%	95.26%	1.88%
TLOβ2 0h (2)	40491332	39256264	96.95%	94.85%	2.10%
TLOβ2 0h (3)	42183358	41022803	97.25%	95.28%	1.97%
TLOβ2 30m (1)	42802076	41567721	97.12%	95.15%	1.97%
TLOβ2 30m (2)	43681928	41728278	95.53%	92.78%	2.75%
TLOβ2 30m (3)	46300450	44913341	97.00%	94.52%	2.48%
TLOγ11 0h (1)	44081170	42822120	97.14%	95.08%	2.07%
TLOγ11 0h (2)	43546274	42250005	97.02%	94.71%	2.32%
TLOγ11 0h (3)	39309424	37832877	96.24%	92.59%	3.66%
TLOγ11 30m (1)	46196636	44814307	97.01%	94.70%	2.31%
TLOγ11 30m (2)	42322228	40862534	96.55%	93.51%	3.04%
TLOγ11 30m (3)	38931274	37096340	95.29%	93.00%	2.29%

Appx. 4 Table 2 Mapping statistics from RNA-seq read alignment

Appendix 5 – Quality statistics for ChIP sequencing

Anti-RNA Polymerase II		
Sample	**Raw reads**	**Q30(%)**
AHY input	42314122	89.97
AHY IP1	7969570	94.73
AHY IP2	47103512	93.32
ΔΔtlo input	54545312	94.02
ΔΔtlo IP1	6971435	95.2
ΔΔtlo IP2	38837500	93.33

Anti-HA		
Sample	**Raw reads**	**Q30(%)**
TLOα1 input	42314122	89.97
TLOα1 IP1	44964216	93.56
TLOα1 IP2	39788364	93.99
TLOβ2 input	49638462	92.35
TLOβ2 IP1	56058652	94.17
TLOβ2 IP2	48983684	92.28
TLOγ11 input	45690664	92.91
TLOγ11 IP1	40087168	92.61
TLOγ11 IP2	41050700	94.52

Appx. 5 Table 1 ChIP-sequencing quality statistics

Appendix 6 – Commands for ChIP-seq data analysis

Quality filtering and adaptor trimming

Quality filtering and Illumina adaptor trimming with performed with the fastp tool (Chen *et al.*, 2018) using the default settings with the following command:

```
fastp –i inputR1.fq.gz –I inpurR2.R2.fq.gz –o outR1.fq.gz –O outR2.fq.gz
```

Alignment

Read files (fq.gz) were aligned to reference with Bowtie2 aligner (Langmead and Salzberg, 2012) and piped into Samtools (Danecek *et al.*, 2021). Then sorted and indexed with Samtools:

```
bowtie2 –k 1 –x bowtie_sorted_reference_file_library –1 reads_1.fq.gz –2 reads_2.fq.gz | samtools view –bS – > alignment.bam
```

```
samtools sort –T temp –O bam –o alignment.sorted.bam alignment.bam
```

```
samtools index alignment.sorted.bam
```

Peak calling

MACS2 (Zhang *et al.*, 2008) was used to call peaks for individual replicates:

```
macs2 callpeak –t IP.alignment.sorted.bam –c input.sorted.bam ––format BAMPE –g 1.4e7 –B ––name samplename ––outdir narrow_peak_calling
```

Bedtools intersect, part of the Bedtools suite (Quinlan and Hall, 2010), was used to determine how many of the peaks in each replicate overlapped with each other. Default setting were used, and the unique (-u) argument was used:

```
bedtools intersect –a rep1.broadPeak –b rep2.broadPeak –u > intersects.bed
```

Investigation of peak locations relative to genes

Bedtools intersect was used to analyse where on the reference genome the peaks were located:

```
bedtools intersect –a reference.gtf –b peaks.narrowPeak > output.bed
```

Converting bedgraph to bigwig

The bedgraph files that were outputted from MACS2 were converted to bigwig files using the BedGRaphToBigWig tool (Kent *et al.*, 2010) in order to visualise the alignments on IGB (Freese *et al.*, 2016). Where chrom.sizes is a .txt file with the chromosome names and sizes listed.

```
bedGraphToBigWig in.bedGraph chrom.sizes out.bw
```