THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Establishing a standardised DNA assembly and biosensor-based screening pipeline for natural product pathway engineering in *Saccharomyces cerevisiae*

Jamie Yam Auxillos

Doctor of Philosophy

Institute of Cell and Molecular Biology
School of Biological Sciences
University of Edinburgh

2020

Dedication

To Lope, Florence, John and Thelma.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification.

Jamie Yam Auxillos
2020

Acknowledgements

[This work was funded by a Gates foundation studentship by the Bill and Melinda Gates foundation through the Global Health program with a continuation in funding supported by the Volkswagen foundation.]

Firstly, I would like to thank Prof. Patrick Cai for the opportunity to do a PhD at his laboratory, specifically on the YeastFab and SynNP projects. I am incredibly grateful for his guidance and support, as well as providing an excellent environment to grow as a researcher. I would also like to thank Prof. Chris French for his invaluable advice on my project and support throughout my PhD, especially during the thesis write up stage.

Many thanks to the Edinburgh Cai lab members, including Roy, Wei, Darek, Katarina, Paulina, Chantal, Andrea, Aileen, Emily, Eva, Daniel, Alba and Erika, who have supported me and made this journey an enjoyable one. In particular, I would like to thank Eva for her invaluable advice and guidance with the SynNP project. Additional thanks to members of the Edinburgh Genome Foundry, including Valentin, Isaac, Anaïs, Hille and Aitor. I would also like to thank all the members of the Manchester Cai lab, including Sally, Reem, Marco, Tao, Jackie, Jun, Jason, Isaac, Jack, Chantal, Daniel and Eva. You have all made the life in Manchester an unforgettable one. Thanks to the members of the Dai lab for all their help and support for the YeastFab collaborative project, in particular, Shuangying, Tianyi, Yakun and Tong.

Many thanks to Prof. Mike Tyers, Almer, Jing, Samuel, Susan, Prof. Gerry Wright, Beth, Grace and Venky for their guidance and help in the collaborative SynNP project. I would also like to thank the Synbiochem group, for helping me with the analysis of viomycin, palmitoleic acid, rebeccamycin and naringenin yeast producer strains. In particular, Mark and Adrian for guidance with the Hamilton system and Cunyu, Kat and Rehana for invaluable help with the chemical analysis of my yeast producer strains. Many thanks to the Suess group at the Technical University of

Darmstadt in Germany. In particular, Beatrix, Flo, Adrien, Cristina, Jeannine and Adam for providing a welcoming environment during my visit and being excellent and supportive collaborators for the (naringenin/viomycin) riboswitch project. Thank you to members of the Micklefield group, in particular, Helen Taylor and Prof. Jason Micklefield, for help with the chemical analysis of rebeccamycin in yeast producer strains. I would also like to thank the Swain group, in particular, Ish, Luis and Nahuel, for their guidance and advice on plate reader experiments and data analysis.

I want to express my gratitude to Dr. Edward Wallace and Rachel, whose guidance and support have kept me going. Many thanks to the members of the Wallace lab, in particular, Laura for helping me proofread a chapter on my thesis, Sam and Rosey, Marah, Weronika, Liz, Flic, Clemence, Abhi and Savvas for their constant support and advice with my thesis. Thanks to members of my PhD committee, Prof. Gary Loake and Prof. Manfred Auer, as well as Dr. Liz Bayne and Dr. Stephen Wallace for advising me on my PhD project. Thank you to members of the graduate school, in particular, Caroline Proctor, Helena and Karen who have helped me during the transition to Manchester and back up to Edinburgh for the write-up stage. Thanks to members of the Rosser lab, in particular, Jess, Matt, Pascoe, Adam, and Mai-Britt for all the help and advice over the years. Many thanks to James for the regular academic consultation.

I would like to thank my family (Lope, Florence, John, uncle Robert, auntie Thelma and tita Bonet) for their constant support over the years. Finally, I would like to thank my father, for his inspiring words.

'Maraming bagay ang napakahirap gawin. Ang ibang bagay kailangan ng tiyaga. Nguni't ang ilan, kailangan mong pagpawisan at minsan konting tenga-tenga lang yan.'

- Lope M. Auxillos

Abstract

Many pharmaceutical drugs currently available on the market are natural products, which are extracted from biological sources (microorganisms or plants). It can prove challenging to maintain a constant supply of these drugs, considering that they are often derived from microorganisms that are difficult to culture or plants, where availability can be strongly impacted by demand for alternative crops or severe weather conditions. Synthetic biology presents a promising and more sustainable option, through the engineering of easily cultured and genetically tractable microorganisms such as *Saccharomyces cerevisiae* to enable their production of pharmaceutical compounds. However, the process of heterologous expression of a bacteria-derived natural product pathway in *S. cerevisiae* can be challenging. Many variables need to be adjusted, such as codon optimisation, gene expression, pathway flux, enzyme variants utilised and the activity of different enzymes within the pathway. In the process of optimisation and screening, a major rate-limiting step is the assembly of diverse pathway variants, which can be time consuming and costly.

The aim of this project was to utilise synthetic biology tools and techniques to engineer yeast strains to produce different therapeutic compounds. This involves the establishment of a pipeline to be able to 1) assemble many different variants of large pathways, and 2) rapidly screen for the production of a bioactive compound and identify producer or high producer strains. In this body of work, I describe the development of YeastFab, a toolkit of standardised and modular biological yeast 'parts' that can be used to hierarchically assemble large natural product pathways that encode for production of therapeutic drugs within the span of a week. To date, hundreds of promoters and terminators have been standardised into the YeastFab format and have been further characterised to determine their activities. This assembly strategy is subsequently applied in the context of a collaborative project, funded by the Bill and Melinda Gates Foundation, for heterologous expression of exemplar pathways from different classes of natural products (non-ribosomal peptides (NRPS), polyketides (PKS), ribosomally

synthesized and post-translationally modified peptides (RiPPs), nucleosides, alkaloids and flavonoids). I have primarily focussed on the heterologous production of viomycin, rebeccamycin and naringenin. Different variants of each pathway have been assembled and screened using a bioassay to detect antibiotic production or using a biosensor specific to the compound of interest. As a proof of principle, I have developed a high throughput assay, utilising an *Escherichia coli* transcription factor-based biosensor to quantify the amount of naringenin produced by each yeast strain, demonstrating comparable quantification to mass spectrometry measurements. Further improvements of naringenin yield are being carried out through chassis optimisation by genome engineering techniques such as SCRaMbLE (Synthetic Chromosome Rearrangement and Modification by LoxPsyn-mediated Evolution), with the use of the *E. coli* biosensor to sift through the population of SCRaMbLEd strains for high producers.

In conclusion, I have developed and demonstrated a pipeline for rapid assembly and screening of natural products biosynthetic pathways in yeast, which can facilitate the development of novel processes for manufacture of many therapeutic compounds.

Lay summary

At present many of the pharmaceutical drugs available are extracted from natural sources, such as plants and different bacteria. When extracted from plants, yields are often low and demand can lead to environmental destruction. In the case of bacteria, many are challenging to grow in large amounts, which can lead to high costs. Chemical manufacturing of these drugs is not always possible, even when the compound structure is known. As such, there has been much interest in exploring alternative methods for the large scale synthesis or extraction of pharmaceutical drugs.

One promising method is the use of baker's yeast as a micro-factory for the mass production of chemicals. This is achieved through the introduction of genes that encode for enzymes that produce these therapeutic drugs. My work has in particular focussed on developing and expanding the tools and approaches we use for engineering the yeast genome, using a genetic toolkit called YeastFab. This toolkit allows us to quickly and easily explore how different combinations and versions of genes that encode for production of compounds can lead to more efficient production within yeast. In my work I have focussed on the production of medically relevant compounds, such as viomycin (an anti-tuberculosis drug) and naringenin (an antioxidant) within yeast. I demonstrate a new approach for screening the amount of naringenin produced by the yeast micro-factory. This method uses a fluorescence based approach, where yeast that produce more naringenin lead to stronger fluorescence, allowing us to select for the best producers. This method can ultimately help us to optimise production and can potentially be further applied for the screening and optimisation of production for other pharmaceutical compounds within yeast.

Contents

List of Figures **xix**

List of Tables **xxiii**

1 Introduction **1**

1.1 Brief history of synthetic biology 1
1.2 Engineering biology 2
1.3 Standardisation of biology 3
1.4 Improvements in DNA assembly techniques 4
1.5 Interdisciplinary nature of synthetic biology 5
1.6 Synthetic biology for cell factory optimisation 5
1.7 Genome engineering methods for cell factory optimisation 7
1.8 Screening for high producers 8

2 Materials and Methods **11**

2.1 DNA, oligonucleotides and stock solutions 11
2.1.1 DNA and oligonucleotides 11
2.1.2 Antibiotic stocks 11
2.1.3 Amino acid stocks 12
2.2 Bacterial plates and liquid media 13
2.2.1 Lysogeny broth (LB) liquid media and agar 13
2.2.2 Super Optimal Broth (SOB media) 13
2.2.3 Mueller Hinton agar 13
2.2.4 Overlay assay agar (Phosphate buffered agar) 14
2.2.5 4x LB media (for the naringenin spent media assay) 14
2.2.6 4x LB buffered media (for the naringenin spent media assay) . . 14
2.3 Yeast plates and liquid media 15
2.3.1 YPD liquid media and agar 15

2.3.2 Preparation of minimal synthetic (SC) dropout media 15
2.3.3 YPD 5'FOA plates 16
2.4 Molecular Biology Techniques 17
2.4.1 Bacterial competent cell preparation 17
2.4.2 Bacterial transformation by heat shock 17
2.4.3 Bacterial plasmid extraction 18
2.4.4 PCR amplification 18
2.4.5 Golden Gate assembly 19
2.4.6 Gibson assembly master mix preparation 19
2.4.7 Yeast competent cell preparation 20
2.4.8 Yeast transformation 20
2.4.9 Yeast genomic DNA extraction 21
2.4.10 Yeast RNA extraction 21
2.4.11 Yeast plasmid extraction 21
2.4.12 Sanger sequencing 21
2.4.13 Whole genome sequencing 22
2.5 Protein analysis 22
2.5.1 Western blot 22
2.6 Promoter characterisation 24
2.6.1 Plate reader growth and fluorescence measurements 24
2.6.2 Normalisation of plate reader data 25
2.7 Antibiotic screening assays 26
2.7.1 Overlay assay with *Bacillus subtilis* or *Micrococcus luteus* 26
2.7.2 Modified overlay assay with *Mycobacterium smegmatis* 26
2.7.3 Antibiotic susceptibility assay 27
2.8 Viomycin chemical analysis 28
2.8.1 Viomycin extraction and screening by LC-MS 28
2.9 Naringenin and *p*-coumaric acid chemical analysis 28
2.9.1 High-throughput chemical analysis of *p*-coumaric acid and naringenin production 28
2.9.2 Naringenin mass spectrometry data analysis 29
2.10 Naringenin spent media assay 30
2.10.1 Spent media assay of naringenin producers - growth curve . . . 30
2.10.2 Spent media assay of naringenin producers - endpoint reading . 30
2.10.3 Interpolation of naringenin production using the spent media assay 31

3 A standard modular cloning toolkit for pathway assembly in yeast 33
3.1 Chapter 3 introduction . . . 33
3.1.1 The YeastFab toolkit . . . 35
3.2 Chapter 3 results . . . 42
3.2.1 Promoter library construction . . . 42
3.2.2 Modification of the HcKan vectors to allow *ccd*B-based screening 49
3.2.3 Modification of the POT vectors to allow *ccd*B-based screening . 52
3.3 Chapter 3 discussion . . . 53

4 High-throughput characterisation of promoters 57
4.1 Chapter 4 introduction . . . 57
4.1.1 Methods for optimising bioproduction . . . 57
4.1.2 Regulation of gene expression . . . 58
4.1.3 Constitutive promoters . . . 59
4.1.4 Regulatable promoters . . . 62
4.1.5 Riboswitch-based regulation of protein production . . . 64
4.2 Chapter 4 results . . . 67
4.2.1 Characterisation of promoter expression profiles over time . . . 67
4.2.2 High-throughput characterisation of promoter expression profiles over time . . . 73
4.2.3 Characterisation of the Z3EV transactivation system . . . 84
4.2.4 Characterisation of the neomycin riboswitch system . . . 90
4.2.5 Domestication of the neomycin riboswitch to YeastFab standards 93
4.3 Chapter 4 discussion . . . 98

5 Constructing and diversifying the viomycin pathway 103
5.1 Chapter 5 introduction . . . 103
5.1.1 A brief history of antibiotic discovery platforms . . . 103
5.1.2 Reinvigorating drug discovery: Using synthetic biology for drug discovery . . . 104
5.1.3 SynNP Project overview . . . 104
5.1.4 Viomycin . . . 107
5.1.5 Viomycin mechanism of action . . . 107
5.1.6 Viomycin biosynthesis . . . 108
5.2 Chapter 5 results . . . 109
5.2.1 Viomycin susceptibility tests) . . . 109
5.2.2 Assembly of the ORF into the HcKan_O acceptor vector . . . 110

5.2.3 Assembly of the full pathway with a set of 'strong promoters' . 112
5.2.4 Combinatorial assembly of viomycin pathway scaffolds with different promoters . . . 121
5.2.5 Minimal viomycin core assembly . . . 130
5.2.6 Analysis of biosynthetic enzyme production by Western blot . . 137
5.3 Chapter 5 discussion . . . 140

6 Construction and high-throughput screening of naringenin producer strains 143
6.1 Chapter 6 introduction . . . 143
6.1.1 Improving the flux of precursor molecules . . . 145
6.1.2 Development of biosensors . . . 147
6.2 Chapter 6 results . . . 151
6.2.1 Domestication of the naringenin biosynthetic enzyme genes to the YeastFab format . . . 151
6.2.2 Transcription unit assembly of the naringenin biosynthetic enzyme genes (POT assembly) . . . 151
6.2.3 Protein expression validation by Western blot . . . 154
6.2.4 Multi-transcription unit assembly of naringenin BSGs . . . 156
6.2.5 Full pathway assembly of the naringenin pathway . . . 158
6.2.6 Characterization of *p*-coumaric acid and naringenin production in JAy078, JAy079, JAy080 and JAy081 strains . . . 161
6.2.7 Assembly of the Nar AT pathway . . . 163
6.2.8 Analysis of the *4CL* enzyme . . . 169
6.2.9 Assembly of a truncated naringenin pathway . . . 172
6.2.10 Establishing an assay for measuring naringenin production . . . 178
6.2.11 Initial screening of the naringenin biosensor . . . 179
6.2.12 Testing for media conditions affecting the naringenin biosensor activity . . . 180
6.2.13 Screening for improvements in production through genomic rearrangements . . . 190
6.3 Chapter 6 discussion . . . 192

7 Discussion 197

Appendix A Primers used in this study 201
A.1 Table of Primers . . . 201

Appendix B Strains used in this study **253**

Appendix C Supplementary figures and tables - for Chapter 3 **259**

Appendix D Supplementary figures and tables - for Chapter 4 **263**

Appendix E Supplementary figures and tables - for Chapter 5 **281**

Appendix F Supplementary figures and tables - for Chapter 6 **293**

Bibliography **301**

List of Figures

3.1 Abstraction hierarchy of biological systems. 34
3.2 Classification of YeastFab parts 36
3.3 Orientation of BsaI recognition sites 37
3.4 Domestication of parts into the YeastFab format (Promoter) 39
3.5 YeastFab assembly pipeline and its expansion 41
3.6 Pipeline for the high-throughput domestication of parts into the Yeast-Fab format 43
3.7 Comparison of promoters that have failed in the PCR and colony PCR steps in in Chr 4 plate 3 and Chr 2 plate 3 47
3.8 The distribution of success rates of each step across all 17 promoter construction plates 48
3.9 Identifying the effects on internal BsaI sites. 49
3.10 Modification of the screening method from red-white based to a positive selection using *ccd*B. 50
3.11 Modification of the screening method from red-white based to a positive selection using *ccd*B. 53

4.1 Diagrammatic representation of the structure of a yeast promoter) . . . 59
4.2 Promoter characterisation pipeline used by Guo et al (2015) 62
4.3 Diagrammatic representation of the GAL system 64
4.4 The Golden Gate assembly strategy for cloning promoters into the reporter vector 70
4.5 Characterisation of a pilot set of 9 promoters 72
4.6 Comparison of relative promoter activities of 74 promoters between two datasets 77
4.7 Characterisation of 'strong' promoters over time 79
4.8 Characterisation of 'moderate strength' promoters over time 80
4.9 Characterisation of 'weak strength' promoters over time 81

4.10 Promoters that show high activity at exponential phase 83
4.11 Promoters that show high activity at stationary phase 84
4.12 Overview of the Z_3EV inducible system 87
4.13 Characterisation of the Z_3EV inducible system 89
4.14 Overview and initial characterisation of the neomycin riboswitch system 92
4.15 Domestication of the neomycin riboswitch to the YeastFab standard . . 94
4.16 Characterisation of the neomycin riboswitch system using the Re2.8 reporter vector 96
4.17 Comparison of neomycin switching factors 98

5.1 A diagrammatic representation of the SynNP project goals 106
5.2 Viomycin biosynthesis 109
5.3 An overview of the viomycin assembly pipeline 111
5.4 Digestion verifications of viomycin POT assemblies 114
5.5 Digestion verifications of viomycin YFass assemblies 117
5.6 Screening of full pathway assemblies (Strong promoter assembly) 119
5.7 Screening of JAy001 using the overlay assay against *B. subtilis* 121
5.8 Combinatorial assembly viomycin BSGs with a pool of promoters of different activities. 122
5.9 Digestion simulation and agarose gel verification of all 20 combinatorial viomycin POT assemblies 124
5.10 Digestion simulation and agarose gel verification of all 4 combinatorial viomycin assemblies in the YFass vector 126
5.11 Screening of viomycin combinatorial assembly clones by colony PCR . . 127
5.12 Screening of the combnatorial viomycin strains using overlay assay with t*B. subtilis* 129
5.13 Assembly of the minimal viomycin core 131
5.14 Screening of minimal viomycin assemblies by colony PCR 135
5.15 Figure 11 136
5.16 Western blot of all 20 viomycin biosynthetic enzymes 140

6.1 A schematic representation of the pathway for naringenin production in *S. cerevisiae* 144
6.2 Mechanism of FdeR-mediated transcription regulation 148
6.3 Overview of the naringenin assembly pipeline and subsequent screening using a biosensor 150
6.4 Verification of naringenin POT assembles (round 1) 154

6.5 Verification protein production of each naringenin biosynthetic enzyme (round 1) . . . 156
6.6 Verification naringenin YFass assemblies (round 1) . . . 158
6.7 Screening of yeast strains containing the naringenin YAC for the presence of all 7 BSGs (round 1) . . . 160
6.8 Quantification of naringenin production by mass spectrometry (round 1) 161
6.9 Quantification of naringenin and *p*-coumaric acid production by mass spectrometry of strains fed with *p*-coumaric acid (round 1) . . . 162
6.10 Verification of POT assemblies (round 2) . . . 165
6.11 Verification of protein production of each naringenin biosynthetic enzyme (round 2) . . . 166
6.12 Verification of YFass assemblies (round 2) . . . 167
6.13 Verification of YFass assemblies (round 2) . . . 168
6.14 Verification of YFass assemblies (round 2) . . . 169
6.15 Comparison of different *4CL* variants . . . 171
6.16 PCR verification of naringenin POTs in the truncated naringenin assembly 174
6.17 Western blot of naringenin *4CL* BSG enzymes. A green asterisk is used to label bands of interest on each lane . . . 175
6.18 Digestion verification of truncated YFass vectors and the digestion simulation . . . 176
6.19 PCR verification of JAy093 . . . 177
6.20 Quantification of naringenin and *p*-coumaric acid production using the mass spectrometer. . . . 178
6.21 Initial testing of the naringenin biosensor . . . 180
6.22 Screening of the biosensor in media with different pH . . . 181
6.23 Screening of the biosensor in different media conditions . . . 183
6.24 Screening of the biosensor in spent yeast media . . . 185
6.25 Screening of the biosensor in the spent yeast media of JAy093 . . . 187
6.26 Biosensor-based analysis of naringenin titers in different producer strains 189
6.27 Biosensor-based analysis of naringenin titers in different diploid producer strains . . . 191

C.1 Scoring of promoters in the PCR and colony PCR step using the PCR gels and colony PCR agarose gels of promoters on Chr 4 plate 3 . . . 260
C.2 Scoring of promoters in the PCR and colony PCR step using the PCR gels and colony PCR agarose gels of promoters on Chr 2 plate 3 . . . 261

D.1 Normalised growth curves of the pilot (9) promoter characterisation . . 264
D.2 Cloning of all 74 promoters into the Re2.8 vector 265
D.3 Growth curve of strong promoters 266
D.4 Growth curve of moderate strength promoters 267
D.5 Growth curve of weak promoters 268
D.6 Verification of Z3EV integration into the SASy31 TRP1 locus 269
D.7 Construction of modified neomycin riboswitches 270

E.1 A disk diffusion assay screening wild type yeast (BY4741) for susceptibility to viomycin 281
E.2 A disk diffusion assay screening *B. subtilis* and *M. smegmatis* for susceptibility to viomycin 282
E.3 Sequencing verification results for all 20 viomycin BSGs domesticated to the YeastFab format 284
E.4 Next generation sequencing of JAy001 for the presence of the full viomycin pathway 286
E.5 Initial screening of combinatorial viomycin clones (with 4 sets of primers) 287
E.6 Screening of for all genes in combinatorial viomycin clones 288
E.7 Screening of 5 clones from combinatorial viomycin assembly by disk diffusion assay 290
E.8 Disk diffusion assay and further chemical analysis of JAy021 and JAy096 291
E.9 Total protein of 20 viomycin POT-containing yeast cell lysates on an SDS-PAGE gel. 292

F.1 Sequencing of domesticated naringenin BSGs-I 294
F.2 Sequencing of domesticated naringenin BSGs-II 295
F.3 Analysis of protein production in POT assemblies (round 1) - Total protein gel 296
F.4 Sequencing verification of JAy081) 296
F.5 Analysis of protein production in POT assemblies (round 2) - Total protein gel 297
F.6 Analysis of protein production in POT assemblies (round 3) - Total protein gel 297
F.7 Sequencing analysis of JAy093 by nanopore sequencing 298

List of Tables

3.1 Table summarising the success rates at every step in the promoter construction pipeline 46

4.1 Promoter information of the 9 promoters and their respective ORF functions 68
4.2 Mutations observed in the 78 promoters upon sequencing verification . 75

5.1 Promoters, terminators and POT vectors used for each viomycin BSG assembly (strong promoter assembly) 113
5.2 Plan for the multi-transcription unit assembly for the viomycin pathway (Strong promoter assembly) 115
5.3 Promoters, terminators and POT vectors used for each viomycin BSG assembly. Each BSG was assembled with a pool of 3 promoters in a combinatorial assembly. 123
5.4 Assembly plan for the multi-transcription unit assembly for the combinatorial viomycin pathway and the calculated size of the final multi-transcription unit insert (bp) 125
5.5 List of promoters used in the pool for the minimal viomycin core assembly 132
5.6 Table of the promoters, terminators and POT vectors used for the minimal viomycin core assembly. Each BSG was assembled with a pool of 3 promoters in a combinatorial assembly. The total size of each viomycin BSG assembly for all 11 genes was calculated (bp). 133
5.7 Assembly plan for the multi-transcription unit assembly for the minimal viomycin pathway and the calculated size of the final multi-transcription unit insert (bp) 134
5.8 Sizes of protein in kDa for each viomycin biosynthetic enzyme. 139

6.1 Naringenin POT assembly scheme (round 1) 153
6.2 Predicted protein sizes of the naringenin biosynthetic enzymes (round 1) 155

6.3 Naringenin YFass assembly scheme (round 1) 157
6.4 Naringenin POT assembly scheme (round 2) 164
6.5 Predicted protein sizes of the naringenin biosynthetic enzymes (round 2) 165
6.6 Naringenin YFass POT assembly scheme (round 2) 167
6.7 Naringenin YFass POT assembly scheme (round 2) 173
6.8 Naringenin YFass assembly design for the truncated naringenin pathway. Labelled in dark green is the YFass using the same POT as N1-trunc YFass 1. 175

A.1 Table of Primers 201

B.1 E coli competent cells used in this study 254
B.2 Table of bacterial used in this study 255
B.3 Table of yeast strains used in this study 256

D.1 26 'strong' promoters and the characterised activity from the Guo et al. (2015) publication. The corresponding ORFs and key functions as described in the Saccaromyces genome database (SGD - https://www.yeastgenome.org) of each of the 26 promoters in this library. 271
D.2 26 'moderate strength' promoters and the characterised activity from the Guo et al. (2015) publication. The corresponding ORFs and key functions as described in the Saccaromyces genome database (SGD - https://www.yeastgenome.org) of each of the 26 promoters in this library. 274
D.3 26 'moderate strength' promoters and the characterised activity from the Guo et al. (2015) publication. The corresponding ORFs and key functions as described in the Saccaromyces genome database (SGD - https://www.yeastgenome.org) of each of the 26 promoters in this library. 277

Chapter 1

Introduction

1.1 Brief history of synthetic biology

Synthetic biology is already beginning to realize its potential, with a wealth of examples showcasing the successful engineering of biological organisms to address real-world problems. Particular highlights include the adaptation of microorganisms for the production of pharmaceutical drugs alongside the generation of biosensors capable of detecting toxins or specific viruses in the field (Paddon *et al.*, 2013; Bereza-Malcolm *et al.*, 2017; Tancharoen *et al.*, 2019). It is a field that has emerged and flourished as a result of multiple breakthroughs in the field of genetic engineering during the second half of the 20th century. The discovery of restriction enzymes (Kelly and Smith, 1970) first allowed the *in vitro* manipulation of DNA and the resulting application in recombinant DNA technology (Jackson, Symons and Berg, 1972; Cohen *et al.*, 1973). In combination with the synthesis of oligonucleotides and discovery of polymerase chain reaction technique (Mullis and Faloona, 1987), all these discoveries have opened up possibilities for the engineering of organisms for applications in biotechnology and medicine. Towards the 1990s, major developments were made in sequencing technologies and associated computational tools leading to the generation of a vast catalogue of genetic information from multiple different organisms. Within this new genomics era, the field of systems biology was born in part out of the large availability of data alongside the aim to investigate the interactions of chemicals and pathways at a systems level. In order to account for the large number of interconnecting pathways and dynamically interacting components, mathematical and computational modelling strategies are employed (Kitano, 2002). In the turn of the century, the vast amount of genetic

information of organisms coupled with an improved understanding of gene regulatory networks paved the way for ambitious design projects with a goal of constructing genetically modified organisms that carry out relatively complex functions of societal value. This resulted in what is currently known as the field of synthetic biology. While systems biology hopes to dissect and explore the intricacies of complex biological systems, synthetic biology takes a more practical approach in a sense that synthetic genetic systems are created and tested within the context of a complex biological system in order to tease apart truths from the resultant changes in phenotype. The heavily used Feynman quote 'What I cannot create, I do not understand' does in a nutshell, provide an accurate representation of the driving motivation of synthetic biology – namely to learn by making.

1.2 Engineering biology

This complexity of biological systems, along with a multitude of unknown or unexpected interactions makes it difficult to precisely engineer for specific purposes. What differentiates synthetic biology from the field of genetic engineering in the 1980s and 1990s is the emphasis applying engineering principles. These include the importance of design and modelling prior to experimentation, as well as the incorporation of concepts such as decoupling, standardisation and abstraction (Endy, 2005). The concept of 'decoupling' involves the fragmentation of a complicated problem into several simpler chunks allowing them to be independently tackled. The application of decoupling in synthetic biology reorients the focus away from the biology of the organism being engineered, and instead, towards the genetic circuit or pathway being designed for the purposes of incorporating a novel or beneficial function. Whilst the decoupling or orthogonality of designed circuits is often only partial in practice, due to the reliance upon intracellular resources and the robustly regulated nature of cells, the underlying principle focusses on the generation of circuits that enable researchers to 'cut through the noise' to achieve specific phenotypic effects.

1.3 Standardisation of biology

Abstraction involves the organisation of biological functions into abstraction hierarchies. It functions as a heuristic to manage the complexity, whereby researchers often specialise at different levels of the abstraction hierarchy while exchanging only the essential information required between levels (Endy, 2005). Standardisation is another core concept applied in synthetic biology, where biological parts (such as promoters, ribosomal binding sites (RBS), open reading frames (ORFs) and terminators) are designed to incorporate standardised features (such as the presence of specific restriction sites) to not only allow for its re-use and interchangeability but also to allow for rapid assembly of these components to form genetic circuits and pathways. Within synthetic biology, a lot of focus and attention has been given to the design and construction of biological systems using standard DNA 'parts'. Two major efforts towards standardisation of biological parts in the field were 1) the development of the BioBrick standard and 2) the organisation and establishment of the international Genetically Engineered Machine competition (iGEM). The utilisation of this parts-based approach, with predefined design specification was demonstrated through the development of the BioBrick standard. These BioBricks form the building blocks for subsequent assemblies of larger and more complex biological circuits through a restriction-ligation based assembly method (Knight, 2003, 2007). The annual, global iGEM competition has helped in particular to increase the adoption of the BioBrick standard, requiring entrants to submit constructed circuits using the BioBrick standard into a shared repository. The increasing number of parts and related characterisation data has led to the establishment of data registries and repositories for parts and devices. Among these are the Registry of Standard Biological Parts (http://parts.igem.org/), the Standard European Vector Architecture database (SEVA-DB) (http://seva.cnb.csic.es/) and the Joint BioEnergy Institute's Inventory of Composable Elements (JBEI-ICE) (https://acs-registry.jbei.org) (Decoene et al., 2018). In addition, the introduction of the Synthetic Biology Open Language (SBOL) visual served as a uniform method for graphically illustrating genetic constructs. SBOL has been widely used for communicating design principles and adopted for computer-aided design (CAD) software (Quinn et al., 2015).

1.4 Improvements in DNA assembly techniques

Whilst a large part of this advancement is due to, or made possible by, rapid decreases in DNA synthesis costs, improvements in downstream cloning technologies has in particular provided synthetic biologists with the capability to more rapidly generate constructs and circuits for subsequent testing. These advancements in cloning techniques can be subdivided into two main categories, endonuclease mediated assembly methods and long-overlap-based methods (Casini *et al.*, 2015). Endonuclease methods were initially used as a method to improve cloning through standardisation. The Biobrick standard uses endonucleases such as EcoRI, SpeI, XbaI and PstI for the standardisation of cloning (Knight, 2007; Shetty, Endy and Knight, 2008). However, several limitations in using this system such as the difficulty in reusing the part in a different plasmid, remaining 'scars' (unwanted segments of DNA remaining as a result of the cloning process) and the requirement to remove internal 'forbidden digestion sites' within the DNA parts can made these tools challenging to use. Type IIs restriction endonucleases provides an alternative, circumventing the problem of 'scars' remaining in DNA sequences following the cloning process (Casini *et al.*, 2015). These restriction endonucleases differ as the recognition site lies outside of the cut site, allowing the user to determine the desired sequence of the overhang. This is the enzyme upon which the popularised Golden Gate assembly method has been based on (Engler, Kandzia and Marillonnet, 2008), and is used for subsequent application in several standardised toolkits for the hierarchical assembly of multiple genes (Sarrion-Perdigones *et al.*, 2011; Weber *et al.*, 2011; Werner *et al.*, 2012; Guo *et al.*, 2015; Lee *et al.*, 2015; Iverson *et al.*, 2016; Moore *et al.*, 2016; Andreou and Nakayama, 2017). On the other hand, long-overlap-based assembly methods rely on the cloning of parts, where each successive part to be cloned possesses homology to the subsequent part, ultimately enabling seamless cloning of multiple parts. One main advantage of these techniques is the lack of a requirement for the removal of forbidden digestion sites. Some examples of these tools include CPEC (circular polymerase extension cloning) (Quan and Tian, 2009), Gibson assembly (Gibson *et al.*, 2009), LCR (Ligase. Cycling reaction) (Kok *et al.*, 2014) and SLiCE (seamless ligation cloning extract method) (Li and Elledge, 2007). The wide range of assembly techniques developed has enabled the assembly of larger segments of DNA, allowing the potential for introducing functionality in microorganisms which was previously unfathomable.

1.5 Interdisciplinary nature of synthetic biology

Synthetic biology has come a long way since the age of recombinant DNA technology and genetic engineering, with major development of tools and techniques enabling rapid assembly of DNA, within a span of weeks. This has been applied, not only for use in biotechnology and medical applications but also for the construction of systems to help answer fundamental biological questions. Additionally, the tools developed in this field has also been applied in the area of synthetic chemistry for the production of novel compounds or synthesis of complex chemicals (Paddon *et al.*, 2013; M. Li *et al.*, 2015; Li *et al.*, 2016; Casini *et al.*, 2018). These tools have also enabled the construction of minimal genomes and synthetic chromosomes (Gibson *et al.*, 2008; Dymond *et al.*, 2011; Annaluru *et al.*, 2014; Hutchison *et al.*, 2016). However, the focus of this body of work will centre on the utilisation of synthetic biology tools for the construction of microfactories for high-value chemicals and compounds.

1.6 Synthetic biology for cell factory optimisation

A significant proportion of pharmaceutical drugs currently available on the market consist of natural products, most of which are extracted from microorganisms or plants (Pickens, Tang and Chooi, 2011). However, this is becoming increasingly unsustainable as demand from low yielding plant sources leads to environmental destruction and the production of a large amount of waste. Additionally, it can become challenging to maintain a constant supply of these medicines when the availability of these drugs varies due to environmental catastrophes, instabilities in agricultural yields and fluctuations in the supply chain. The extraction of these compounds from microorganisms which naturally produce these drugs pose its own set of problems; namely that many of these organisms are challenging to grow in a laboratory environment, making industrial-scale production unfeasible (Pickens, Tang and Chooi, 2011). An alternative is the refactoring and metabolic engineering of industrially tractable microorganisms to function as a microfactory for the production of high value compounds (Krivoruchko and Nielsen, 2015; Luo *et al.*, 2015). The ability to enhance yields and decrease costs of production and extraction, as well as reducing end waste products, has encouraged the adaptation of bio-production by industry (Kung *et al.*, 2018). The process of refactoring a biosynthetic pathway is, however, often laborious and time

consuming. Beginning with the identification of biosynthetic enzyme genes or gene clusters, many bioinformatic tools, such as anti-SMASH (antibiotics and Secondary Metabolite Analysis Shell) (Medema et al., 2011) and NPsearcher (Li et al., 2009) have been developed for mining of genomes for sequences responsible for the production of these high value compounds. These biosynthetic enzyme genes are either directly amplified from the original producer or synthesised *de novo* by a commercial vendor and subsequently cloned into one or multiple different chassis to screen for the production of the compound or its intermediates. Multiple iterations of pathway re-design and modification, alteration of the host primary metabolism for improved precursor flux and alleviation of bottlenecks is carried out in order to obtain a cell factory able to produce moderate-to-high titres of the compound off interest. Several research groups have utilised *in silico* modelling or computer-aided design (CAD) tools for the optimisation of yeast metabolic flux in order to improve the production of the desired compound (Asadollahi et al., 2009; Brochado et al., 2010). Nevertheless, much progress has been made in this field, with many research groups and companies successfully engineering microorganisms to produce a variety of biopharmaceutical compounds, including antitumor drugs (Qu *et al.*, 2015; Li *et al.*, 2018), antibiotics (Gidijala *et al.*, 2009; Siewers *et al.*, 2009; Awan *et al.*, 2017; Shomar *et al.*, 2018) and antioxidants (Yan, Kohli and Koffas, 2005; Koopman *et al.*, 2012; M. Li *et al.*, 2015; Li *et al.*, 2016; Rodriguez *et al.*, 2017) within the past decade. The improvement of DNA assembly tools previously mentioned has enabled the assembly of larger pathways over the years. Of note is the recent breakthrough of by Galanie *et al.*, in their successful assembly and production of the opiods - thebaine and hydrocodone, which consist of 21 and 23 enzymes respectively (Galanie *et al.*, 2015). This is the first demonstration of the assembly of a pathway consisting of 21-23 biosynthetic enzyme genes and the successful production of the desired opioid chemical compounds from the action of these enzymes in yeast.

Synthetic biology tools and concepts have been applied to the field of metabolic engineering. In particular, standardisation and modular DNA assembly resulting in the adaptation of the 'plug-and-play' concept in the field. This has enabled the rapid prototyping of enzyme combinations, not only for the refactoring and optimisation of pathway expression but also for the combinatorial diversification using tailoring enzymes. These methods result in the construction of large libraries of genetically diverse strains that can be subsequently screened for successfully refactored variants.

Shao *et al* employed such a 'plug-and-play' scaffold for the successful expression and optimisation of a silent biosynthetic gene cluster encoding spectinabilin (Shao *et al.*, 2013). Cummings *et al.* has demonstrated the successful application of these tools for the diversification of polyketide scaffold structures in *Eschericchia coli*through the combinatorial assembly with a library of tailoring enzymes resulting in the identification of novel polyketide structures (Cummings *et al.*, 2019). Similarly, Yim *et al.* employed this method in *Streptomyces coelicolor* for the identification of novel variants of the glycopeptide antibiotic (Yim *et al.*, 2016). At the same time, increasing abundance and annotation of metagenome data, computational tools and improved prediction of protein activity have provided a much wider range of enzymes that can be screened for activity, increasing the substrate for diversification.

1.7 Genome engineering methods for cell factory optimisation

Ongoing improvements in DNA assembly and genome engineering techniques will provide the tools to explore the effect of a greater number of modifications for both the pathway and the chassis genome on pathway refactoring and optimisation efforts. The effect of promoter engineering has been demonstrated by Liu *et al.* using a recombinase system for the shuffling of promoters cloned upstream of β-carotene and violacein strains, altering expression profiles of biosynthetic enzyme genes (Liu *et al.*, 2018). Mitchel 2015 similarly modified the β-carotene pathway through promoter engineering, however through the combinatorial assembly of promoters of different strengths using Golden Gate assembly (Mitchell *et al.*, 2015). Barbieri *et al.* utilised eukaryotic Multiplex Automated Genome Engineering (eMAGE) to modify the promoter, coding sequence and terminator of the β-carotene pathway to optimise its yields (Barbieri *et al.*, 2017). On a genomic level, Bao *et al* have used a method called CRISPR–Cas9- and homology-directed-repair- (HDR) assisted genome-scale engineering (CHAnGE), which enabled rapid construction and screening of mutant libraries, to isolate a strain with improved furfural tolerance (Bao *et al.*, 2018). Similarly, researchers have utilised a synthetic yeast strain, which contains multiple LoxP sites on its chromosome, enabling chromosomal rearrangements by cre-EBD (inversions, duplication, translocation) upon induction with estradiol, to investigate modifications on the yeast chassis for compound production. Using this system, Blount *et al.* obtained a strain capable of increasing

the expression of pathways, such as violacein and penicillin, on 2μ plasmids, ultimately doubling their compound yields (Blount *et al.*, 2018). Similarly, Liu *et al.* applied this method for the optimization of β-carotene and violacein pathway titers and obtained strains with two-fold (500 μg/L) and ten-fold higher production (10 mg/L) respectively, compared to non-optimised strains (Liu *et al.*, 2018).

1.8 Screening for high producers

Optimisation of a heterologous pathway in a chassis organism by promoter or genome engineering often results in large libraries of genetically diverse strains. Similarly, diversification of a successfully refactored scaffold pathway for novel drug discovery also produces a large population of strains that require subsequent screening. Their subsequent screening becomes a large bottleneck in the chassis optimisation pipeline as their individual screening can become labour intensive and expensive as analytical methods rely on medium to low throughput methods such as mass spectrometry and chromatographic techniques. Despite ongoing improvements in these technologies, their throughput lags behind that of strain construction, necessitating an intermediary method to enrich the fraction of productive strains. Currently, methods for screening large libraries of genetically diverse strains often rely on a colorimetric output of the compound (Barbieri *et al.*, 2017; Liu *et al.*, 2018). Biosensors provide a promising alternative for the screening of strains producing a compound that do not present a visual indication of yield. Most commonly used biosensors are riboswitches and transcription factor-based (TF) biosensors, which are systems that are regulated by ligand binding, either to an aptamer or a transcription factor, resulting in an alteration of reporter gene expression (e.g. fluorescence). Riboswitches can be constructed using aptamers designed for many compound using the SELEX (systematic evolution of ligands by exponential enrichment) technique (Stoltenburg, Nikolaus and Strehlitz, 2012; Lauridsen *et al.*, 2018). However, the *in vitro* selected aptamers with high affinity for a compound may not necessarily function when coupled with a reporter element and can require further engineering (Groher and Suess, 2014). Despite this, several aptamers regulated by small molecules (Wallis *et al.*, 1995; Weigand *et al.*, 2008; Stoltenburg, Nikolaus and Strehlitz, 2012; Jang *et al.*, 2017; Xiu *et al.*, 2017; Groher *et al.*, 2018) have been isolated by SELEX and successfully engineered as riboswitches.

Li *et al.* showcased the strength of TF biosensors, employing a TF based biosensor for malonyl-CoA, a necessary component for polyketide and flavonoid biosynthesis. They demonstrated its use for screening a cDNA overexpression library to isolate yeast strains with improved malonyl-CoA production (Figure 5A) (S. Li *et al.*, 2015). Similarly, Skjoedt *et al.* adapted cis, cis-muconic acid (CCM) and naringenin binding transcriptional activators from prokaryotic LysR-type transcriptional regulators (LTTRs) for yeast functionality (Figure 5A) (Skjoedt *et al.*, 2016). However, when the produced compound does not accumulate in the cell, this can cause inaccuracies in the measurement due to fluctuations in its intracellular concentration. Jang *et al.* presented an alternative via co-culturing a naringenin producer strain with an *E. coli* strain containing a naringenin riboswitch biosensor (Jang *et al.*, 2017; Xiu *et al.*, 2017). This variant of co-culturing, however, may not serve as a method amenable for ultra-high-throughput screening due to the need for strain deconvolution prior to screening.

Coupling biosensors with genetically diverse strains enables rapid, initial screening for high production strains (Abatemarco *et al.*, 2017). The population of yeast cells exhibiting high yields exemplified by high fluorescence can then be carried forward for subsequent rounds of genome engineering for further diversification, whilst continuing to select for gains in yield. The identified highest producers can then be further analysed by mass spectrometry and chromatographic techniques for the precise quantification of the compound of interest. In addition, whole genome and RNA sequencing can be carried out to elucidate the causal links between the genotypic alterations and the associated phenotypes, to help establish design principles for engineering of related products.

In this thesis, I will discuss the construction and optimisation of a modular toolkit, YeastFab (Chapter 3). Parts domesticated as part of the toolkit have been characterised and described in Chapter 4. The YeastFab toolkit is then applied the assembly and expression of the viomycin (Chapter 5) and naringenin pathways (Chapter 6) in yeast. Finally, in Chapter 6, I discuss the development and optimisation of a spent media assay for the detection of naringenin production by yeast strains.

Chapter 2

Materials and Methods

2.1 DNA, oligonucleotides and stock solutions

2.1.1 DNA and oligonucleotides

Oligonucleotides used in this study were synthesised by Integrated DNA Technologies (Leuven, Belgium). The exceptions were oligonucleotides used to amplify promoters for the YeastFab library. These oligonucleotides synthesised and delivered lyophilised in 96 well PCR plates by the Wuxi Qinglan Biotech Co. Ltd (Yixing City, China).

De novo DNA synthesis of the biosynthetic enzyme genes were made by Gen9 (Cambridge, Massachusetts) from 2014-2016. Provider of DNA synthesis services was switched to Twist and Wuxi Qinglan Biotech Co. Ltd (Yixing City, China) due to the acquisition of the company by Gingko Bioworks (Boston, Massachusetts).

2.1.2 Antibiotic stocks

Ampicillin stocks were prepared by dissolving 100 mg/ml of ampicillin sodium salt (Sigma Aldrich cat no. A9518-5G) in sterile water and filter sterilised using a 0.22 μM syringe filter unit (Merck Millipore cat no. SLGP033RS). These were aliquoted in 1.5 ml Eppendorf tubes and stored at -20°C.

Kanamycin stocks were prepared by dissolving 50 mg/ml of kanamycin sulphate (Sigma Aldrich cat no. 10106801001) in sterile water and filter sterilised using a 0.22 µM syringe filter unit (Merck Millipore cat no. SLGP033RS). These were aliquoted in 1.5 ml Eppendorf tubes and stored at -20°C.

Chloramphenicol stocks were prepared by dissolving 25 mg/ml of chloramphenicol (Sigma Aldrich cat no. C0378-25g) in 100% ethanol and filter sterilised using a 0.22 µM syringe filter unit (Merck Millipore cat no. SLGP033RS). These were aliquoted in 1.5 ml Eppendorf tubes and stored at -20°C.

Viomycin stocks were prepared by dissolving 10 mg of viomycin (Tocris cat. no. 3787) in 162 µl of sterile water to obtain a stock concentration of 70 mM (Maximum solubility concentration of viomycin is 75 mM).

2.1.3 Amino acid stocks

Amino acid stocks in Edinburgh were prepared by Aileen Greig. Amino acid stocks in Manchester were prepared by Dr. Eva Garcia Ruiz.

100 mM Histidine stock was prepared by mixing 21 g of L-Histidine (Sigma-Aldrich cat no. H8000) in 1 litre of distilled water, aliquoted in small bottles and sterilised by autoclaving.

100 mM Leucine stock was prepared by mixing 13 g of L-Leucine (Sigma-Aldrich cat no. L8000) in 1 litre of distilled water, aliquoted in small bottles and sterilised by autoclaving.

40 mM Tryptophan stock was prepared by mixing 8 g of L-Tryptophan (Sigma-Aldrich cat no. T0254) in 1 litre of distilled water, aliquoted in small bottles and sterilised by autoclaving. Typtophan stocks were stored in °C.

20 mM Uracil stock was prepared by mixing 2.24 g of Uracil (Sigma-Aldrich cat no. U1128) in 1 litre of distilled water, aliquoted in small bottles and sterilised by autoclaving.

100 mM Lysine stock was prepared by mixing 18.3 g of L-Lysine (Sigma-Aldrich cat no. L5501) in 1 litre of distilled water, aliquoted in small bottles and sterilised by autoclaving.

16 mM Adenine stock was prepared by mixing 3 g of Adenine (Sigma-Aldrich cat no. A2786) in 1 litre of distilled water, aliquoted in small bottles and sterilised by autoclaving.

2.2 Bacterial plates and liquid media

2.2.1 Lysogeny broth (LB) liquid media and agar

LB media was prepared by adding the following components; 5g of tryptophan (BD diagnostic cat no. 211705), 2.5g of yeast extract (BD diagnostic cat no. 212710), 5g of sodium chloride (Fisher Chemical S/3160/60), 7.5g of agar [only for preparing agar plates] (Formedium cat no. AGA04) and distilled water up to a total volume of 500 ml. LB media were sterilised by autoclaving.

For LB plates requiring antibiotic selection, the agar was microwaved then cooled to 50°C. The appropriate amount of antibiotics from pre-prepared stock solutions were added to a final concentration of 100 μg/ml for ampicillin, 50 μg/ml for kanamycin or 25 μg/ml for chloramphenicol. For LB liquid media requiring antibiotic selection, the appropriate amount of antibiotics to a final concentration of 100 μg/ml for ampicillin, 50 μg/ml for kanamycin or 25 μg/ml for chloramphenicol and stored at 4°C.

2.2.2 Super Optimal Broth (SOB media)

Super Optimal Broth (SOB) media was prepared by adding 20 g of Bacto tryptone (Gibco cat no. 211705), 5 g of yeast extract (BD diagnostic cat no. 212710), 0.58 g of sodium chloride, and 0.19 g of potassium chloride. These components were dissolved in 500 ml of distilled water and autoclaved to sterilise.

2.2.3 Mueller Hinton agar

Mueller Hinton plates were prepared using Mueller Hinton agar powder purchased from Sigma (cat no. 70191-500G) and prepared according to the manufacturer's instructions.

2.2.4 Overlay assay agar (Phosphate buffered agar)

The 75 mM phosphate buffered agar overlay was prepared by mixing 0.8 g of Nutrient broth no. 4 (Merck cat no. 03856-500G), 0.4 g of sodium phosphate monobasic monohydrate (Merck cat no. 71507-250G), 0.65 g of sodium phosphate dibasic (Merck cat no. S7907-100G) and sterile water to 100 ml. The pH was measured and adjusted to pH7. 0.7 g of agar (Formedium cat no. AGA04) was added then autoclaved. This protocol was developed and optimised by Dr. Grace Yim and Elizabeth Culp from the Wright lab (Prof. Gerard Wright) at McMaster University.

2.2.5 4x LB media (for the naringenin spent media assay)

4x LB medium was prepared by dissolving 100g of Bacto-tryptone (Gibco cat no. 211705), 50g of yeast extract (BD diagnostic cat no. 212710), 100g of sodium chloride in a total of 250 ml of sterile water.

Two versions of this 4x LB supplement was tested - one without pH adjustment and one with the pH adjusted 7. In both cases, the media was filter sterilised using a single use bottle tip filtration system (Thermo cat no. 291-3320). 4x chloramphenicol was added for selective growth of bacteria containing a chloramphenicol resistance cassette and stored in 4°C.

2.2.6 4x LB buffered media (for the naringenin spent media assay)

4x LB medium was prepared by dissolving 100g of Bacto-tryptone (Gibco cat no. 211705), 50g of yeast extract (BD diagnostic cat no. 212710), 100g of sodium chloride, 3.2 g of sodium phosphate monobasic monohydrate (Merck cat no. 71507-250G) and 5.2g of sodium phosphate dibasic (Merck cat no. S7907-100G) in a total of 250 ml of sterile water.

Two versions of this 4x LB supplement was tested - one without pH adjustment and one with the pH adjusted 7. In both cases, the media was filter sterilised using a single use bottle tip filtration system (Thermo cat no. 291-3320). 4x chloramphenicol was added for selective growth of bacteria containing a chloramphenicol resistance cassette

and stored in 4°C. The optimal formulation which was used for the quantification of naringenin production in yeast producer stains is the 4x LB buffered media adjusted to a pH of 7.

2.3 Yeast plates and liquid media

2.3.1 YPD liquid media and agar

YPD liquid media and YPD agar consist of following components; 1% of yeast extract (5g) (Formedium cat no. YEA03), 2% of peptone (10g) (Formedium cat no. PEP03), 2% of dextrose (10g) (Fisher Chemical cat no. G/0500/53) and 2% of agar (10g) [only for preparing agar plates] (Formedium cat no. AGA04) and distilled water up to a total volume of 500 ml. YPD liquid media and YPD agar was sterilised by autoclaving.

2.3.2 Preparation of minimal synthetic (SC) dropout media

Synthetic complete (SC) dropout media (Sc-3 and Sc-6) was prepared as 2x stocks, adding 0.85 g of yeast nitrogen base without amino acids and ammonium sulfate (BD Difco cat no. DF0335-15-9) and and 1.0 g of Sc Kaiser dropout mix (For Sc-6 dropout media - Formedium SC Kaiser, Multiple Dropout -Ade, -His, -Leu, -Lys, -Trp, -Ura) (For Sc-3 dropout media- Formedium SC Kaiser, Triple Dropout -His, -Leu -Ura for Sc-His-Leu-Ura media) and 2.5 g of ammonium sulfate (Fisher Chemical cat no. A/6480/65). The mixture was dissolved completely in 250 ml of distilled water and sterilised by autoclaving.

1x Sc liquid media was made from 2x Sc (Sc-3 or Sc-6) dropout media with 25 ml of 40% w/v glucose, additional amino acids where required and distilled water to a final volume of 500 ml. Amino acid supplement solutions were added based on the desired dropout media required.

For Sc-His single dropout media, 10 ml of 100 mM leucine and 5 ml of 20 mM uracil amino acid stocks were added to the 2x Sc-3 dropout media. For Sc-Leu single dropout media, 1.5 ml of 100 mM histidne and 5 ml of 20 mM uracil amino acid

stocks were added to the 2x Sc-3 dropout media. For Sc-Ura media, 1.5 ml of 100 mM histidine and 10 ml of 100 mM leucine amino acid stocks were added to the 2x Sc-3 dropout media.

For Sc-His -Leu double dropout media, 5 ml of 20 mM uracil amino acid stock was added to the 2x Sc-3 dropout media. For Sc-His -Ura double dropout media, 10 ml of 100 mM leucine amino acid stock was added to the 2x Sc-3 dropout media.

For Sc-His -Leu -Trp triple dropout media, 5 ml of 16 mM adenine, 5 ml of 100 mM lysine and 5 ml of 20 mM uracil amino acid stocks were added to the 2x Sc-6 dropout media. For Sc-His -Leu -Lys triple dropout media, 5 ml of 16 mM adenine, 5 ml of 40 mM tryptophan and 5 ml of 20 mM uracil amino acid stocks were added to the 2x Sc-6 dropout media.

For Sc complete media, 1.5 ml of 100 mM histidne, 10 ml of 100 mM leucine and 5 ml of 20 mM uracil amino acid stocks were added to the 2x Sc-3 dropout media.

For Sc agar, 4% w/v agar (2x) was prepared by mixing 10g in 250 ml of distilled water and sterilised by autoclaving. 25 ml of 40% w/v glucose and the appropriate amino acids as described aboe were added the sterile 2x Sc media to obtain the desired 2x Sc dropout media. The sterile 2x agar was microwaved until completely melted and the 2x Sc dropout media was mixed and poured on sterile petri dishes.

For spent media assay of naringenin producers, strains were grown in Sc-His -Ura MSG media with a pH between 6.5-7. This media was prepared as 2x stocks, adding 0.85 g of yeast nitrogen base without amino acids and ammonium sulfate (BD Difco cat no. DF0335-15-9) and and 1.0 g of Sc Kaiser dropout mix (For Sc-3 dropout media-Formedium SC Kaiser, Triple Dropout -His, -Leu -Ura for Sc-His-Leu-Ura media) and 0.5 g of L-Glutamic acid monosodium salt monohydrate (MSG) (Sigma-Aldrich cat no. 49621). The mixture was dissolved completely in 250 ml of distilled water and the pH was adjusted to a pH between 6.5-6.8. The media was filter sterilised using a single use bottle tip filtration system (Thermo cat no. 291-3320).

2.3.3 YPD 5'FOA plates

Counterselection by 5-fluoroorotic acid (5-FOA) medium (1 mg/ml) was prepared by adding 0.5 g of 5-FOA (Formedium cat no. 5FOA10) to 250 ml of 2x YPD media,

microwaved for 30 seconds then incubated in a water bath set at 50°C for 1 hour. During this time, 230 ml of 2x agar (4% w/v agar) was microwaved until completely melted. The 250 ml of 2x yeast media with the completely dissolved 5-FOA was added to 4% w/v agar, mixed and poured on petri dishes.

2.4 Molecular Biology Techniques

2.4.1 Bacterial competent cell preparation

SOB media and CCMB buffer were prepared 2 days prior to the bacterial competent cell preparation. CCMB buffer was prepared by adding 5.9 g of calcium chloride dihydrate (Sigma-Aldrich cat no. C3306-100G), 2 g of manganese chloride tetrahydrate (Sigma-Aldrich cat no. 221279), 1 g of magnesium chloride hexahydrate (Sigma-Aldrich cat no. M9272) and 500 µl of 5M potassium acetate (Sigma-Aldrich cat no. 95843-100ML-F).

E. coli cells (DH5α, Mach1) were inoculated in 5 ml of LB media and grown overnight in a 37°C shaking incubator. The next day, 500 ml of SOB was inoculated with the 5 ml overnight *E. coli* culture and the OD_{600} of the culture was measured using a spectrophotometer (expected OD_{600} to be less than 0.2). The culture was grown in a 37°C shaking incubator until an OD_{600} of 0.4. Cells were then harvested and washed three times with 50 ml CCMB buffer. Cells were resuspended in 20 ml of CCMB buffer and incubated on ice for 1 hour then aliquoted into PCR tubes for storage in -80°C.

2.4.2 Bacterial transformation by heat shock

Plasmid DNA was transformed into competent *E. coli* cells by heat shock (Gingold, E. 1985). Competent cells prepared according to the protocol described above (50 µl aliquots) were thawed on ice. Once thawed, DNA was added to the competent cells in sterile conditions (maximum volume of DNA was of 10% of the competent cell volume). The competent cells were then incubated on ice for 10 minutes then transferred to a 42°C water bath for 45 seconds then incubated on ice for 5 minutes. In sterile conditions, 500 µl of LB media was added to each tube and grown in a 37°C

incubator for 1 hour. The cells were pelleted at 12,000g for 30 seconds and 300 µl of the supernatant was removed. The remaining supernatant was used to resuspend the cells. 100 µl of the transformation reaction was plated onto selective media and spread evenly using glass beads (different volumes can be plated according to the amount of plasmids transformed). The plates were dried then incubated at 37°C overnight.

2.4.3 Bacterial plasmid extraction

5 ml cultures of the strain containing the plasmid of interested were grown in selective media (LB with antibiotic selection) overnight in a rotating cell culture roller drum (Eppendorf). On the next day, plasmids were extracted from bacterial cultures using the QIAprep Spin Miniprep kit (Qiagen cat no. 27106) according the manufacturer's protocol. Purified plasmids were quantified using the Nanodrop Spectrophotometer (Thermo Fisher cat no. ND-2000)

2.4.4 PCR amplification

Different polymerases were used for the amplification of DNA depending on the purpose. All PCR amplification carried out in this study utilised the PCR thermocycling protocol as defined by the manufacturer.

The high-throughput amplification of promoter parts from the yeast genome for Chapter 3 utilised the ExTaq DNA polymerase (Takara cat no. RR001CC). The reaction was prepared as defined by the manufacturer but scaled down to 10 µl reactions to account for the cost and scale of the project.

The amplification of biosynthetic enzyme genes in Chapters 5 and 6 were using PrimeStar DNA polymerase (Takara cat no. R010B). The PCR reaction was prepared as defined by the manufacturer's protocol for 50 µl samples.

Colony PCR verification was carried out using the GoTaq DNA polymerase master mix (Promega cat no. M3008). The PCR reaction was prepared as defined by the manufacturer's protocol, however scaled down to 10 µl reactions

2.4.5 Golden Gate assembly

Golden Gate assembly was carried out according to the protocol described in Garcia-Ruiz et al. 2018.

For Bsa1 Golden Gate assemblies, the following components were mixed in a PCR tube on ice; 1 µl of 10x T4 ligase buffer (NEB cat no. M0202M), 0.05 µl of T4 ligase (2,000,000 U/ml) (NEB cat no. M0202M), 0.5 µl of BsaI (cat no. R0535L), 1 µl of vector (50 ng), 1 µof insert (1:1 equimolar ratio of insert to vector) and distilled water to a total volume of 10 µl. The reaction was incubating in the following thermocycling reaction 37°C for 5 minutes, then 30 cycles of 37°C for 5 mins and 16°C for 10 minutes. This is followed by a 30 min incubation at 16°C and a 1 hour incubation at 37°C. The enzyme and ligase are inactivated at 50°C for 5 minutes then 80°C for 10 minutes and finally the program is held indefinitely at 4°C.

Esp31 Golden Gate assemblies, follow a similar protocol as that for BsaI described above except that BsaI is replaced with Esp3I (NEB cat no. R0734L).

2.4.6 Gibson assembly master mix preparation

The Gibson assembly used in this study is based on the protocol published by Gibson et al. 2009. 6 ml of 5x ISO buffer was prepared by adding 3 ml of 1M Tris-HCl pH 7.5 (Invitrogen cat no. 15567-027), 300 µl of 1 M MgCl2 (Sigma cat no. M-1028), 60 µl of 100 mM dGTP (NEB cat no. N0446S), 60 µl of 100 mM dATP (NEB cat no. N0446S), 60 µl of 100 mM dTTP (NEB cat no. N0446S), 60 µl of 100 mM dCTP (NEB cat no. N0446S), 300 µl of 1M DTT (Sigma cat no. D9779), 1.5 g of PEG-8000 (USB cat no. 19959) and 300 µl of 100 mM NAD (Sigma cat no. N1511). The 5x ISO buffer was aliquoted into PCR tubes (40 µl in each) and can be stored in -20°C.

Large batches of the Gibson master mix were prepared and quality control tested for their efficiency before use to prevent pipetting error issues. On ice, the following volumes of each component was mixed together and subsequently aliquoted into PCR tubes. For 1 Gibson assembly reaction, 4 µl of 5x ISO buffer, 0.008 µl of T5 exonuclease (Epicenter cat no. T5E4111K, 10 Uµl), 0.25 µl of Phusion polymerase (NEB cat no. F-530L, 2U/µl), 2 µl of Taq ligase (NEB cat no. M0208L, 40 U/µl) and 8.74 µl of distilled water were mixed together. This can be scaled up according to the number

of samples to prepare. Using this master mix, equimolar amounts of purified DNA fragments (100 ng of vector backbone) was added to a total volume of 20 µl the reaction. The sample was incubated in a PCR thermocycler at 50°C for 1 hour then held at 4°C. The reaction was immediately transformed into *E. coli* competent cells and plated in selective media.

2.4.7 Yeast competent cell preparation

5 ml of yeast media (with appropriate selection) was inoculated with the a single colony of the yeast strain and grown overnight in a cell culture roller drum (Eppendorf) in an incubator set at 30°C. The overnight culture was used to re-inoculate 50 ml of fresh media to a starting OD_{600} of 0.2 and grown in a shaking incubator set at 30°C to an OD_{600} of 0.8. Cells were harvested at 3,000 rpm for 5 minutes, then washed with 10 ml of sterile water. Cells were pelleted by centrifugation at 3,000 rpm for 5 minutes then washed with 10 ml of 0.1 M LiOAc. Cells were pelleted by centrifugation at 3,000 rpm for 5 minutes then the supernatant was discarded. 1 ml of 0.1 M LioAC was used to resuspend the cell pellet.

For frozen competent cells, cells were resuspended with 1 ml of frozen competent cell solution containing 5% v/v glycerol and 10% v/v DMSO in 0.1 M LioAC. This protocol is based on the published method by Gietz et al. 2007.

2.4.8 Yeast transformation

Herring sperm DNA (Promega cat no. D1811) was heat-denatured at 100°C for 10 minutes. Linear DNA (for *in vivo* homologous recombination)/plasmid DNA was transformed into yeast by combining the following components in a 1.5 ml eppendorf tube; 36 µl of 1M LiOAc, 25 µl heat denatured herring sperm DNA (Promega cat no. D1811), 240 µl of 44% PEG 3,350, 36 µl of DMSO and transforming DNA/plasmid. The mixture was vortexed. 50 µl of competent yeast cells were added on top, mixed by inverting several times then incubated at 30°C for 30 minutes. The mixture was transferred to a water bath set at 42°C for 15 minutes then centrifuged for 30 seconds at 10,000 rpm. 200 µl of the supernatant was removed and the remaining liquid used to resuspend the pellet. The cell suspension was plated on selective media using glass

beads and grown for 2-3 days in a 30°C incubator. This protocol is based on the published method by Gietz et al. 2002.

2.4.9 Yeast genomic DNA extraction

Yeast genomic DNA extraction was carried out using the Masterpure Yeast DNA purification kit (Lucigen cat no. MPY80200) according to the manufacturer's protocol. The concentration of the extracted genomic DNA was measured using a nanodrop and verified by running it on an agarose gel prior to its use for further experimentation.

2.4.10 Yeast RNA extraction

Yeast genomic DNA extraction was carried out using the Masterpure Yeast RNA purification kit (Lucigen cat no. MPY03100) according to the manufacturer's protocol. The concentration of the extracted RNA was measured using a nanodrop and verified by running it on an 2% w/v agarose gel prior to its use for further experimentation.

2.4.11 Yeast plasmid extraction

Plasmid DNA from yeast was extracted using the Zymoprep Yeast plasmid miniprep kits according to the manufacturer's protocol (Zymo cat no. D2004). Fresh yeast cultures grown for 20 hours were used for plasmid extraction. Due to the low amount of extracted plasmid, 5 µg of the extracted plasmid was transformed into *E. coli* cells for propagation

2.4.12 Sanger sequencing

The work for Chapter 3, 4 and 5 were carried out in Edinburgh where plasmid DNA and PCR products were sequenced by the Edinburgh Genomics sequencing facility. Sanger sequencing was performed on the DNA using BigDye Terminator v3.1 cycle sequencing technology (Applied Biosystems). The BigDye reaction was carried out prior to sending the samples for sequencing with Edinburgh Genomics. The reaction

included 2 µl of 5x sequencing buffer, 0.64 µl of the primer at a concentration of 10 µM), 2 µl of BigDye, 200-300 ng of DNA and distilled water to a fnal volume of 10 µl. Thermal cycling settings were carried out according to the recommendations of Edinburgh Genomics.

Majority of the work for Chapter 6 (Naringenin) was carried out in Manchester, where plasmid DNA and PCR products were sequenced by a sequencing facility (Source Bioscience). Plasmid DNA (5µl of a concentration of 100 ng/µl) or PCR products (5 µl of a concentration of 10 ng/µl) along with the primers at a concentration of 3.2 pmol/µl (5 µl) was sent to the Source Bioscience sequencing facility.

2.4.13 Whole genome sequencing

NGS sequencing was carried out by the Sanger Institute. The genomic DNA extraction protocol was based on recommendation by the Sanger Institute and Dr. Daniel Schindler. Genomic DNA extraction was carried out using the MasterPure Yeast genomic DNA extraction kit followed by a phenol/chloroform extraction to isolate high quality DNA for NGS sequencing. A slicone gel was used to allow better phase separation in the phenol/chloroform extraction as described by Mukhopadhyay et al. 1993. The concentration and quality of the extracted genomic DNA was measured using a nanodrop and by Qubit prior to sample shipping to the Sanger institute. Analysis of sequencing data was done by Dr. Daniel Schindler using a previously developed analysis pipeline at the Cai lab and visualised using the Integrative Genomics Viewer (IGV) visualisation tool (Broad Institute). Two strains (JAy093 and JAy104) were also sequenced and analysed using the Nanopore MinION sequencer (Oxford Nanopore) by Dr. Daniel Schindler and Marco Monti.

2.5 Protein analysis

2.5.1 Western blot

The total protein was extracted in 2 growth phases, at exponential phase and late exponential to stationary phase. 5 ml of Sc-Ura media was inoculated with the strains and grown overnight at 30°C. The cells were used to inoculate 15 ml of fresh Sc-Ura

media to an OD_{600} of 0.2. Cells were harvested at an OD_{600} of 0.8-1 (5 ml or 5 ODs) and OD_{600} 8-10 (500 µl or 5 ODs). The total protein of yeast strains of interest was extracted based on the protocol by Zhang and colleagues (Zhang et al. 2011) with the following exceptions. The cells were boiled in 50 µl of SDS-PAGE sample buffer.

3 µl of total protein extract was mixed with 3 µl of a solution of bromophenol blue and 10% v/v glycerol. 4 µl of the mixture was loaded into a 4-15% gradient Mini-Protean TGX Stain-Free Protein gel (15 well) (Biorad cat. no4. 568086) with 2 µl of Precision Plus Protein WesternC standards (Biorad cat no. 1610376). The gel was run at 185V for 15 minutes.

The agarose gel was removed from the plastic cast and imaged on the Chemidoc MP Imaging system (Biorad). The total protein was transferred to a nitrocellulose membrane using a semi-dry system (Trans-Blot Turbo – Biorad) using nitrocellulose transfer packs according to the manufacturer's protocol (Biorad cat no. 1704159). The blots were blocked overnight in 5% w/v milk in PBS (Sigma cat no. P4417-50TAB) with 0.05% Tween-20 (Sigma - P9416-50ML) in a cold room on a rocker. The membrane was subsequently incubated with the primary antibody (Monoclonal ANTI-FLAG M2 antibody produced in mouse – Sigma – F1804-50UG) in PBS with 0.05% v/v Tween-20 at an antibody dilution of 1:1000. This step was carried out in a cold room on a rocker for 1 hour. The membrane was washed PBS with 0.05% v/v Tween-20 for 5 minutes and repeated two more times. The membrane was incubated with the secondary antibody solution containing Anti-Mouse IgG (whole molecule) – Peroxidase antibody produced in rabbit (Sigma cat no.A9044-2ML) at a ratio of 1:100,000 and the Precision Protein StrepTacin-HRP conjugate (Biorad cat no. 1610381) at the dilution of 1:10,000. This was carried out in room temperature on a rocker for 1 hour. The membrane was washed PBS with 0.05% Tween-20 for 5 minutes and repeated two more times. For the detection of chemiluminescence of proteins of interest bound with a HRP secondary antibody, the membrane prepared using the Clarity Western ECL substrates kit (Biorad cat no. 1705061) according to the manufacturer's protocol and imaged using the Chemidoc MP Imaging system (Biorad) under the 'chemiluminescence' setting in the 'blots' category. The membrane was also imaged under the 'stain-free' setting in the 'blots' category for total protein normalisation.

2.6 Promoter characterisation

2.6.1 Plate reader growth and fluorescence measurements

For each of the strains in the initial screening of a pilot plate of 9 constitutive promoters, two clones were grown for each strain in one microtitre plate measurement experiment. Each of the strains containing two clones for each strain were screened for their growth over 24 hours and fluorescence and repeated four times over four separate days (experimental replicates n=4). This was carried out to account for differences in growth and promoter activity between experiments.

As for the high-throughput characterisation of promoters, the 76 promoter reporter strains were divided into 2 plates and screened for their fluorescence activity using a plate reader. Division of the strains into 2 plates was to allow the screening of clones for each strain to allow investigation of the difference between clones in one microtitre plate measurement experiment. The number of clones could not be increased due to limited time and availability of the plate reader. Each of the two plates containing two clones for each strain were screened for their growth over 24 hours and fluorescence and repeated three times over three separate days (experimental triplicate n=3). This was carried out to account for differences in growth and promoter activity between experiments.

100 µl of Sc-Ura media was pipetted to each well of a 2.2 ml 96 deep well plate (Star Lab cat no. E2896-1810) using a reagent reservoir (Argos Technologies cat no. B3125-100) and a multichannel pipette. The plate was covered with a gas permeable seal (Fisher cat no. 10642674) and grown for 24 hours in a shaking incubator (Infors) set at 30°C and 850 rpm. 1 ml of fresh Sc-Ura media was added to each well of the deep well plate. 200 µl of the culture was transferred to to a 96 well clear microtitre plate (Thermo cat no. 612F96). The plate and lid used in these plate reader experiments were optimised by the Swain lab (Iseabail Farquhar, Ivan Clark, Dr. Luis Montano) for low evaporation of the cultures. The OD_{600} was measured using a plate reader (BioTek Synergy H1M). This was to verify that the OD_{600} between 0.18-0.22. 50 µl of each culture on the deep well plate was added to a 96 well black microtiter plate (Greiner cat no. 655906) then 150 µl of fresh Sc-Ura media into each well and carefully resuspended with a multichannel pipette. The absorbance at 600nm, YFP at 500/541

(excitation/emission) and mCherry at 579/616 (excitation/emission) were measured every 10 minutes.

All measurements obtained by the BioTek Synergy plate reader were exported to an excel spreadsheet file from the Gen5 (version 2.0) software. The blank corrected fluorescence measurements (YFP and mCherry) were extracted from the results excel spreadsheet and used for subsequent data analysis.

2.6.2 Normalisation of plate reader data

Due to different lag times between experimental runs (experimental triplicates - ran over separate days), a normalisation was carried out to shift the data sets across three experimental replicates based on the shortest lag time. This was first done by calculating the mid-exponential phase time point (peak dy/dx) of the strains and for each experimental replicate. The time point of fastest growth rate for each strain was calculated by first applying non-linear regression, using a logistic growth equation (Prism8 statistical software). This provided the time point of mid-exponential phase time point for each strain. This was used to align all the data sets across three experimental replicates. The calculation and normalisation carried out on the growth curves was used to shift the YFP and mCherry curves accordingly.

The normalised YFP fluorescence was calculated by dividing the YFP measurements by the mCherry measurements at each time point for each experimental triplicate of each strain. The mean normalised YFP fluorescence (YFP/mCherry) was calculated and plotted against time, with error bars plotted using the standard error of the mean (SEM). This analysis was carried out for the characterisation of the promoter characterisation using Prism 8 data analysis software.

2.7 Antibiotic screening assays

2.7.1 Overlay assay with *Bacillus subtilis* or *Micrococcus luteus*

The overlay assay protocol was developed by Elizabeth Culp and Dr. Grace Yim from the Wright laboratory at McMaster University.

Yeast strain containing the biosynthetic pathway was inoculated in 5 ml of yeast selective media and grown overnight in a rotating cell culture roller drum (Eppendorf) set at 30°C. The culture was serially diluted to $10^{-}5$-$10^{-}6$ and 100 µl was plated on yeast selective media (with inducer) using glass beads and grown for 2 days in a 30°C incubator. This was to obtain a single colonies of the yeast strains.

The test organism (*B. subtilis* or *M. luteus*) was streaked out on LB agar (without selection) and grown overnight at 30°C. A pre-culture of the test organism was set up by inoculating 5 ml of LB media (without selection) with the test organism and grown overnight in a rotating cell culture roller drum (Eppendorf) at 30°C.

The overlay assay agar/phosphate buffered agar was microwaved for 1 minute to completely melt. 8 ml of the agar was transferred to a sterile 15 ml centrifuge tube (Sarstedt cat no. 62.554.001) and kept warm in a water bath set at 50°C (for at least 15 mins). On a gel casting levelling table, the height of the surface was adjusted to ensure that it was levelled. Plates containing single colonies of the yeast strains containing the biosynthetic pathway were placed on the levelling table. The warm overlay agar was then inoculated with 40 µl (1:200 dilution) of the test organism, briefly vortexed and poured on top of the yeast plate. The overlay agar is evenly spreading over the entire surface and left to dry for 15 minutes then incubated overnight at 30°C.

2.7.2 Modified overlay assay with *Mycobacterium smegmatis*

The above protocol was modified to account for differences in the growth of *M. smegmatis*.

Yeast strain containing the biosynthetic pathway was inoculated in 5 ml of yeast selective media and grown overnight in a rotating cell culture roller drum (Eppendorf)

set at 30°C. The culture was serially diluted to $10^{-}5$-$10^{-}6$ and 100 µl was plated on yeast selective media (with inducer) using glass beads and grown for 2 days in a 30°C incubator. This was to obtain a single colonies of the yeast strains.

The test organism *M. smegmatis* was streaked out on LB agar (without selection) and grown for 3-4 days at 37°C. 300 µl of PBS was transferred to an eppendorf tube, then using a tip, *M. smegmatis* colonies were scraped from the plate and resuspended in PBS.

The overlay assay agar/phosphate buffered agar was microwaved for 1 minute to completely melt. 8 ml of the agar was transferred to a sterile 15 ml centrifuge tube (Sarstedt cat no. 62.554.001) and kept warm in a water bath set at 50°C (for at least 15 mins). On a gel casting levelling table, the height of the surface was adjusted to ensure that it was levelled. Plates containing single colonies of the yeast strains containing the biosynthetic pathway were placed on the levelling table. The warm overlay agar (without the *M. smegmatis* suspension) was poured on top of the yeast plate. The overlay agar is evenly spreading over the entire surface and left to dry for 15 minutes. A sterile swab was inoculated with the *M. smegmatis* PBS solution and used to streak a lawn on the solidified overlay agar. The plate was then incubated for 3-4 days at 30°C.

2.7.3 Antibiotic susceptibility assay

Test organisms (*B. subtilis*, *M. luteus* or *M. smegmatis*) were inoculated in 5 ml of LB media and grown overnight in a rotating cell culture roller drum (Eppendorf) in a 30°C incubator. *M. smegmatis* was for 3-4 days at 37°C.

A sterile swab was inoculated with the test organism (*B. subtilis*, *M. luteus* or *M. smegmatis* in the overnight liquid culture and used to streak a lawn on the solidified overlay agar. 5 µl dilutions of antibiotics were prepared and 5 µl of each antibiotic dilution is spotted on the plate and left to dry for 10-15 mins. The plate was then incubated for overnight at 30°C (3-4 days at 37°C for *M. smegmatis*.

2.8 Viomycin chemical analysis

2.8.1 Viomycin extraction and screening by LC-MS

All extraction and chemical analysis of producer strains were carried out by Dr. Venkateswarlu Yarlagadda from the Wright lab at McMaster University. Yeast strains were grown in 3 ml Sc -His-Leu dropout media containing 20% w/v glucose for 24 hours at 30°C. The cells were harvested by centrifugation and subsequently seeded in 100 ml fresh Sc -His-Leu dropout media containing 20% 2 w/v glucose and grown at 30C for 72 hours. The culture was centrifuged to separate the yeast cells from the spent medium.

Pelleted yeast cells were resuspended in 5 ml of 1:1 water/methanol solution and subsequently lysed with glass beads by vortexing for 5 minutes. The cell suspension was flash frozen with liquid nitrogen then centrifuged. The cell lysate was extracted with a triple solvent solution of dichloromethane, water and methanol in a 1:1:1 ratio. The fractions extracted were dried using a Genevac evaporator and subsequently analysed by LC-MS.

The spent medium was extracted using a triple solvent solution of dichloromethane, water and methanol in a 1:1:1 ratio. The extracted fractions were dried using a Genevac evaporator and also analysed by LC-MS.

The antibacterial activity of the extracted fractions (both of the cell pellet and the spent medium) was also tested on an agar plate containing a lawn of tester organisms

2.9 Naringenin and *p*-coumaric acid chemical analysis

2.9.1 High-throughput chemical analysis of *p*-coumaric acid and naringenin production

This naringenin chemical analysis protocol was developed by researchers at the synthetic biology centre (SYNBIOCHEM) at the Manchester Institute of Biotechnology.

Extraction was carried on the SYNBIOCHEM Hamilton liquid handler and the chemical analysis was carried out by Dr. Cunyu Yan using the Waters triple quadrupole mass spectrometer (Waters Tqs-HTP).

This protocol for sample preparation for mass spectrometry was obtained from Dr. Mark Dunstan and altered for yeast growth and extraction. Yeast strains were inoculated in biological triplicates on a 96 deep well plate with 100 µl of Sc-His-Ura media, sealed with a gas permeable seal and grown overnight in a shaking incubator set at 850 rpm in 30°C. The culture was split into two, one for the uninduced condition and one for the induced condition. For the truncated pathway growth and analysis, fresh media containing 150 µM of *p*-coumaric acid was added to the uninduced samples and fresh media containing 150 µM of *p*-coumaric acid and 1µM of β-estradiol was added to the induced samples. For the full pathway growth and analysis, fresh media was added to the uninduced samples and fresh media containing 1µM of β-estradiol was added to the induced samples. The plate was sealed with a gas permeable film and grown overnight in a shaking incubator set at 850 rpm in 30°C. The next day, the plate was processed using the pre-set script written by Dr. Mark Dunstan for flavonoid extraction. This protocol begins with the isolation of a sample of the culture for endpoint OD_{600} reading on the plate reader. This was followed by the centrifugation of the deep well plate at 3000 rpm for 10 minutes. Followed by the transfer of 100 µl of the supernatant to a fresh microtiter plate containing 100 µl of 100% methanol followed by a resuspension step. This was followed by a 1:2000 dilution of the supernatant:methanol mixture witth 10% methanol. The final plate was transferred to a Waters microtiter plate for subsequent analysis on the Waters triple quadrupole mass spectrometer (Waters Tqs-HTP) carried out by Dr. Cunyu Yan

2.9.2 Naringenin mass spectrometry data analysis

Mg/l was calculated from the measured nM. The measured nM was divided by 1,000,000,000 to convert nL to L. The value was multiplied by its molecular weight then multiplied by 1000 to convert grams to mg. The value was multiplied by 4000 to account for the dilution carried out before sample quantification on the triple quadrupole mass spectrometer (Waters Tqs-HTP). Calculation of the amount (mg/l) was carried out by Dr. Cunyu Yan. I then plotted the data on a graph using a statistical analysis software, Prism 8.

2.10 Naringenin spent media assay

2.10.1 Spent media assay of naringenin producers - growth curve

Yeast producer strains were inoculated in 100 µl of Sc-His -Ura MSG media (pH 6.5) in a 96 deep well plate and grown overnight in a shaking incubator set at 850 rpm at 30°C. After 24 hours of growth, the strains were subcultured into 600 µl of Sc-His -Ura MSG media (pH 6.5) containing 150 µM of *p*-coumaric acid for the uninduced condition and 150 µM of *p*-coumaric acid and 1 µM of estradiol for the induced condition. The cells were grown overnight in a shaking incubator set at 850 rpm at 30°C. The next day, the yeast cells on the 96 deep well plate were pelleted and 300 µl of the spent media was transferred to a 96 well (1ml) deep well plate. 100 µl of 4x LB (buffered) was added to each well and inoculated with the naringenin *E. coli* biosensor strain to a starting OD_{600} of 0.2. A standard curve was run using the negative control yeast spent media spiked with dfferent concentrations of naringenin (1 µM-250 µM of naringenin dissolved in DMSO). The 150 µl of cells were aliquotted into a sterile black microtiter plate (Thermo cat no. 612F96) and grown in a BioTek plate reader for 24 hours, with OD_{600} and RFP measurements taken every 5 mins.

2.10.2 Spent media assay of naringenin producers - endpoint reading

Yeast producer strains were inoculated in 100 µl of fresh media in a 96 deep well plate and grown overnight in a shaking incubator set at 850 rpm at 30°C. After 24 hours of growth, the srains were subcultuerd into 600 µl of fresh media containing 150 µM of *p*-coumaric acid for the uninduced condition and 150 µM of *p*-coumaric acid and 1 µM of estradiol for the induced condition. The cells were grown overnight in a shaking incubator set at 850 rpm at 30°C. The next day, the yeast cells on the 96 deep well plate were pelleted and 300 µl of the spent media was transferred to a 96 well (1 ml) deep well plate. 100 µl of 4x LB (buffered) was added to each well and inoculated with the naringenin *E. coli* biosensor strain to a starting OD_{600} of 0.2. A standard curve was run using the negative control yeast spent media spiked with dfferent concentrations of naringenin (1 µM-250 µM of naringenin dissolved in DMSO).

The cells were grown overnight in a shaking incubator set at 850 rpm at 30°C. After 24 hours, the endpoint OD_{600} and RFP were measured using a BioTek plate reader.

2.10.3 Interpolation of naringenin production using the spent media assay

A four-parameter dose-response curve was generated, using the Prism8 statistical software, from the RFP/OD measurements taken on the plate reader for each strain including the concentration curve. Using the dose-response cure it was then possible to interpolate the concentration of naringenin based upon the observed fluorescence for different strains.

Chapter 3

A standard modular cloning toolkit for pathway assembly in yeast

3.1 Chapter 3 introduction

Synthetic biology aims to utilise engineering principles for designing and constructing novel biological systems (parts, circuits and organisms). However, the complexity and unpredictable nature of these biological systems make engineering biology a challenge. To address this, synthetic biologists have turned to the engineering concepts of abstraction, standardisation and modularisation, along with bioinformatics and computational modelling tools to simplify the process of engineering biological systems.

Many yeast synthetic biology tools are built on a framework of abstraction, whereby hierarchies of biological systems consist of DNA that form parts that build circuits or pathways which ultimately control systems (Figure 3.1) (Endy, 2005). The categorisation of these biological parts into defined and standardised classifications simplifies engineering efforts and their modularisation allows reusability for the construction of other systems.

In this chapter, I introduce work carried out as part of a collaboration with researchers at the Dai laboratory at the Shenzhen Institutes of Advanced Technology for the standardisation of yeast biological parts into a modularised Golden Gate assembly format (Engler et al., 2008). Within this collaboration, I have standardised

and modularised hundreds of yeast promoters from the yeast chromosome II, IV and V by cloning them into our parts accepting vector, HcKan_P, following design specifications of the YeastFab toolkit (a Golden-Gate assembly-based toolkit). Cloning or domestication of these promoters into the YeastFab standard allows its usability in both characterisation efforts to understand parts functionality in multiple different contexts, and its utilisation in construction efforts for the engineering of *S. cerevisiae* as a microfactory for the production of high-value compounds (Guo et al., 2015). Our collaborators in the Dai laboratory have focused on the domestication of promoters from the yeast chromosomes VI, VII, VIII and IX.

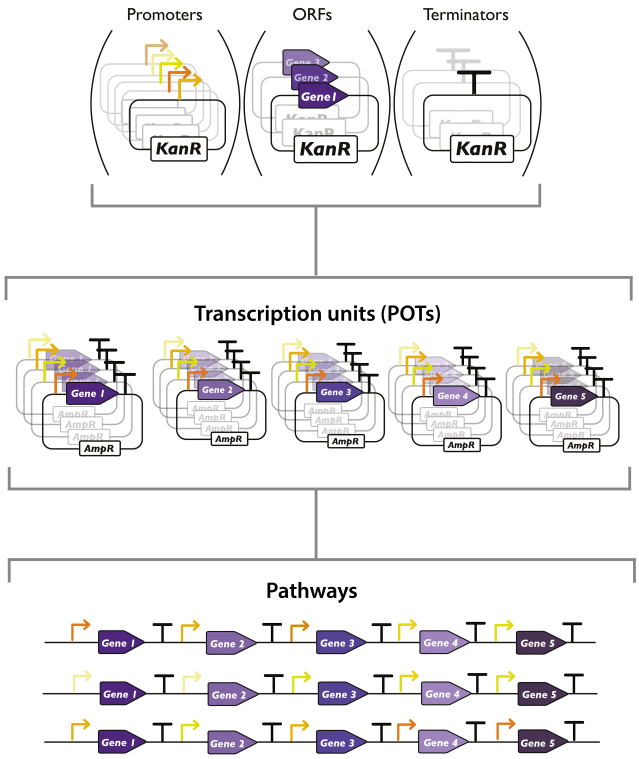


Figure 3.1

Abstraction hierarchy of biological systems. Using this concept, biological parts can be standardised and modularised to allow re usability and interchangeability when constructing more complex circuits or pathways.

3.1.1 The YeastFab toolkit

The YeastFab assembly method is an efficient strategy for the assembly of transcriptional units for expression in *S. cerevisiae*. Biological parts (promoters, CDS and terminators) are domesticated to the YeastFab standard and can be ssubsequently used for the assembly of pathways in a hierarchical fashion using Golden Gate cloning (Guo et al., 2015).

This process of standardisation and domestication into the YeastFab format begins with the identification of 'parts' from the yeast genome. To this end, the sequences of the parts were mined from the yeast genome database (https://www.yeastgenome.org/) as part of the computational tool "Genome Carver" (Scher et al., 2014). Within the YeastFab standard, a promoter is defined as the yeast genomic DNA sequence 500 bp upstream of the ATG of a coding sequence (Figure 3.2A) or the sequence upstream of the ATG up to the adjacent gene boundary (Figure 3.2B). Priority is given to whichever is the shorter sequence to avoid including sequences from the adjacent coding sequence. A coding sequence (CDS) is defined as the sequence from the start codon to the stop codon. A terminator is defined as the 200 bp sequence immediately downstream of the stop codon (Figure 3.2A) or the sequence downstream of the stop codon up to the adjacent gene boundary (Figure 3.2B). Similar to the promoter annotation, priority is given to the shorter sequence to avoid including sequences from the adjacent coding sequence.

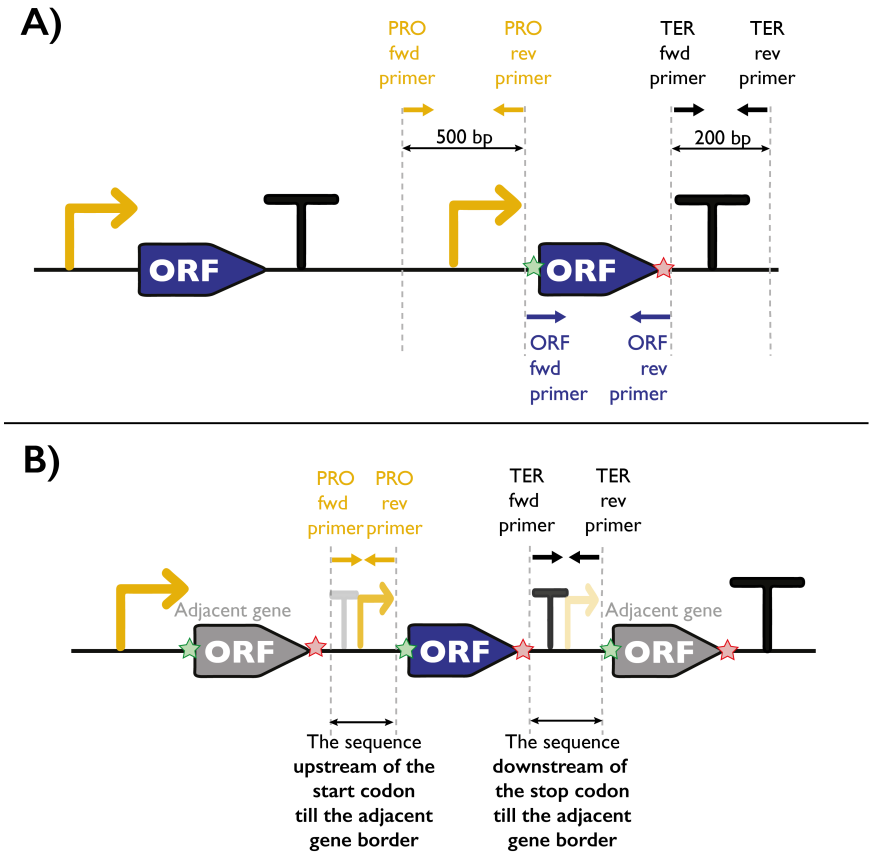


Figure 3.2

A) The promoter sequences are defined as the 500 bp preceediing the start codon and termianator sequences are defined as the 200 bp following the stop codon. B) The sequences are truncated if they overlap with the adjacent gene border

As part of the computational tool, primers were generated to incorporate BsaI restriction sites and appropriate 4 bp overhangs. YeastFab uses type IIs restriction sites that cut outside of their recognition site, revealing 4 bp overhangs (prefix and suffix) and allowing directional cloning. The primers were designed to incorporate an inward facing BsaI recognition site, placing the 4 bp overhang immediately next to the 'part' (Figure 3.3A). Therefore, upon digestion 4 bp overhangs are produced and separate the BsaI recognition sites from the digested part. The first 4 bp overhang upstream of every 'part' is called the 'prefix'. The second 4 bp overhang downstream of the 'part' is the 'suffix'. A unique prefix and suffix are used for each of the three parts to enable subsequent assembly of transcriptional units. GGCT is the prefix and

GATG is the suffix of all promoters, GATG is the prefix and TAGC is the suffix for all ORFs and TAGC is the prefix and CCTC is the suffix of all terminators (Figure 3.3B).

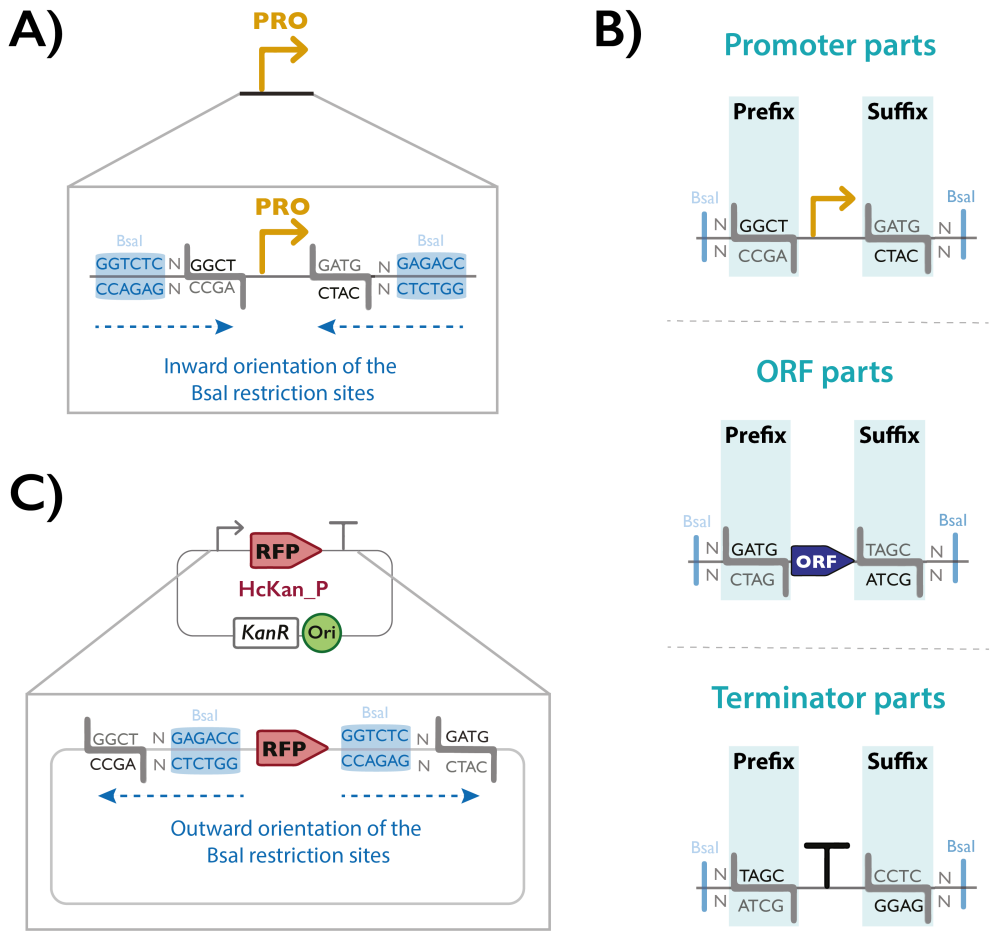


Figure 3.3

A) Orientation of BsaI recognition sites and their corresponding prefix and suffix flanking each promoter part. B) The prefix and suffix flanking each biological part (Promoter, ORF and Terminator). C) Orientation of the BsaI recognition sites and their corresponding prefix and suffix on the HcKan_P vector. These are compatible with the prefix and suffix of the promoter part.

Similarly, parts-accepting vectors (HcKan_P for promoters, HcKan_O for ORFs and HcKan_T for terminators) were designed to contain a bacterial promoter and terminator driving the expression of RFP, which is flanked by BsaI and BsmBI recognition sites that produce the same overhangs. The BsaI sites, which are used for parts domestication, are designed with an outwards orientation relative to the RFP

(Figure 3.3C). Upon digestion with BsaI for cloning a promoter into the HcKan_P vector, RFP is digested out revealing the 4 bp overhangs. Similarly, digestion of the promoter amplicon with BsaI reveals 4 bp overhangs. The overhangs of the promoter part are compatible with the HcKan_P vector allowing their annealing. The ligase in the Golden Gate assembly reaction joins the two parts. The orientations of the BsaI restriction sites on the part and acceptor vector were designed such that no BsaI recognition sites would be incorporated into the part-HcKan assembled vector, resulting in a non-reversible insertion (Figure 3.4), with alternative ligation products (eg. RFP cassette re-ligating into the HcKan vector) resulting in the regeneration of the BsaI site allowing subsequent re-digestion to occur. When the Golden Gate reaction is transformed into *E. coli*, cells containing the original HcKan_P vector containing the RFP cassette result in a red colony. The desired promoter part successfully cloned into the HcKan_P vector appear white as the RFP cassette has been removed and replaced with the promoter (red-white screening). Subsequent assembly of transcription units or POTs rely on the BsmB1/Esp31 restriction enzyme to digest the parts out of the parts accepting vector through BsmB1/Esp31 sites on the HcKan vectors flanking the part.

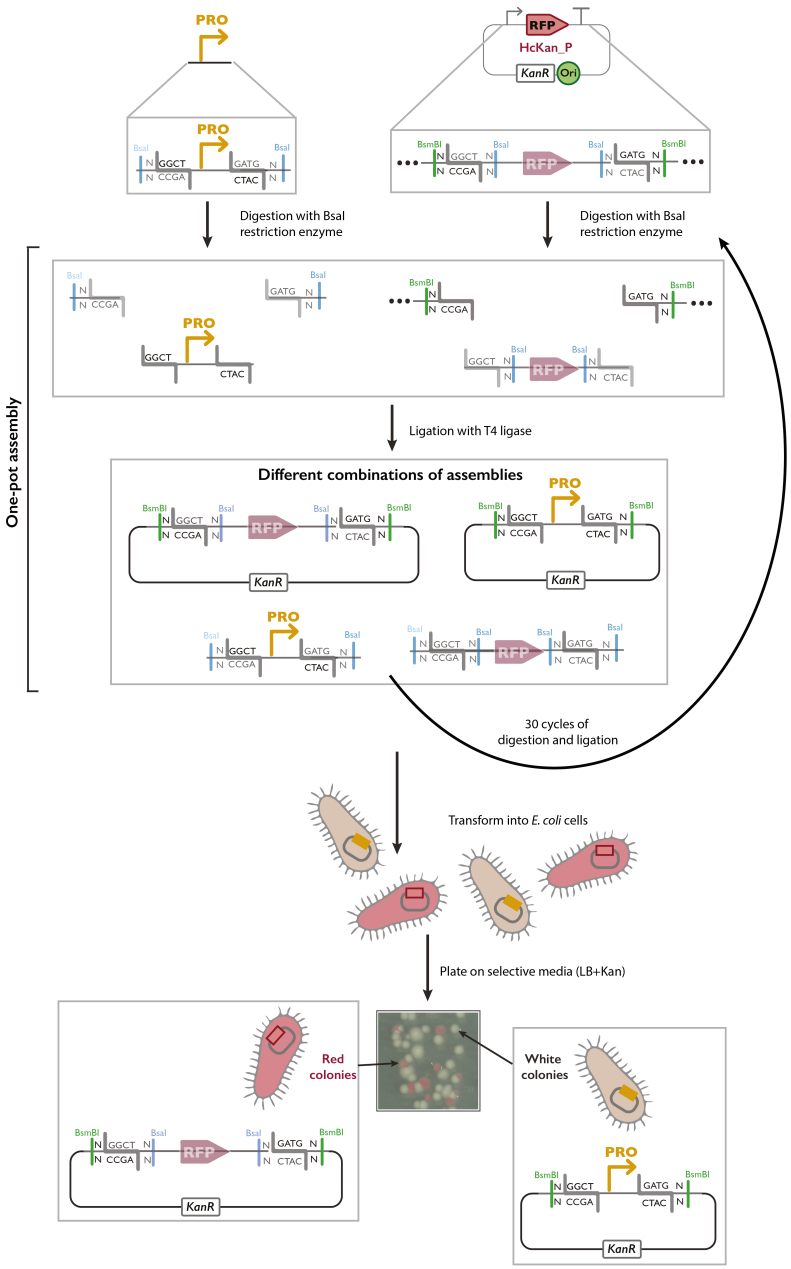


Figure 3.4

Golden gate assembly is used to domesticate parts, in this case a promoter, into the parts accepting vector. The parts accepting vector contain a bacterial promoter and terminator driving the expression of RFP, which is flanked by BsaI and BsmBI recognition sites that produce the same overhangs compatible with the part to be domesticated. Golden Gate ligation is a one-pot reaction, whereby a promoter and the parts accepting vector are incubated with BsaI restriction enzyme and T4 ligase. The digestion with BsaI reveals the prefix and suffix flanking the promoter part and separates the RFP cassette from the HcKan_P acceptor vector. Four possible conformations can result from the ligation of these parts however only one results in the loss of all BsaI recognition sites. The assembly reaction is transformed into *E. coli* cells and subsequently analysed by red-white screening. White colonies are chosen as they have replaced the RFP cassette with a promoter part.

Following their domestication into their respective HcKan vectors, these parts can be subsequently used for transcription unit assembly, also by Golden Gate assembly with BsmB1/Esp31 (Figure 3.5A). Digestion of the domesticated part with Esp3I releases the parts (a promoter, ORF and terminator) from their donor parts accepting vectors (HcKan_P, HcKan_O and HcKan_T respectively) revealing their corresponding prefix and suffix overhangs. The prefix and suffix of all the parts were designed to allow the assembly of the 'parts' in the order of promoter-ORF-terminator (POT) into a POT acceptor vector. This is a yeast shuttle vector with an bacteria promoter and terminator driving the expression of an RFP gene, flanked by BsmBI/Esp31 and BsaI sites. Esp3I digestion releases the RFP cassette, revealing the 4 bp overhangs compatible with the promoter's prefix and the terminator's suffix for transcription unit assembly. The replacement of the RFP cassette with a promoter-ORF-terminator cassette into the POT vector results in white colonies when the Golden Gate assembly is transformed into *E. coli*. The maintenance of the RFP cassette in the POT vector and subsequent transformation into *E. coli* result in red colonies (red-white screening). It must be noted that different POT vectors must be used depending on the size of the full pathway (Figure 3.5B). Digestion with BsaI releases the POTs for subsequent assembly of multiple transcription units into the YFass vectors by Golden Gate cloning (Figure 3.5C). The assembly of multiple transcription units is dependent on another set of 4 bp overhangs, with different combinations in each of the 11 POT acceptor vectors available in the YeastFab system. These allow the assembly of up to 6 transcription units. Therefore, the selection of which POT vector to assemble with each transcription unit, which pathway assembly will rely on, must be decided on before carrying out the assembly of the transcription units.

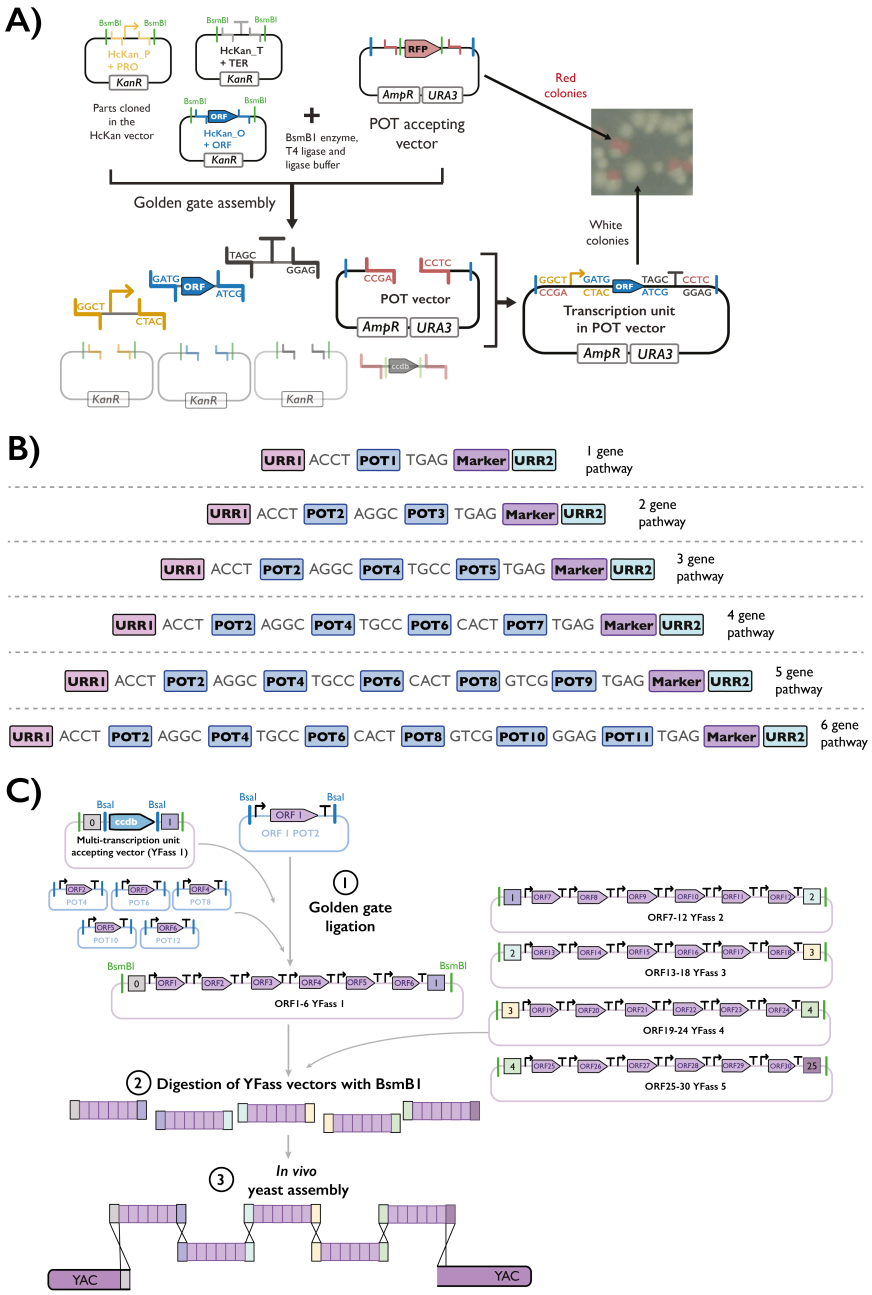


Figure 3.5

A) Transcription units are assembled in the same fashion as shown in Figure 3.4. POT accepting vectors contain a bacterial promoter and terminator driving the expression of RFP, which is flanked by BsaI and BsmBI recognition sites that produce the same overhangs compatible with the promoter-ORF-termrinator cassette. However in this case, BsmBI/Esp31 is used to digest the parts out of the parts accepting vector and in the same reaction, digest the POT vector, separating the RFP cassette from the POT vector. Golden Gate assembly is used to clone the parts into the POT vector in the order of promoter-ORF-terminator. Red-white screening is used to isolate clones with the correct assembly. B) The POTs that need to be used for the assembly of multiple gene pathways. Adapted from (Garcia-Ruiz et al., 2018). C) The expansion of YeastFab to allow the assembly of up to 30 gene pathways.

The design and construction of the multi-transcription unit acceptor vectors were previously carried out by Dr. Wei Liu to expand the YeastFab assembly system to allow the assembly of large pathways of up to 30 genes (Figure 3.5C). Five vectors were constructed containing a *ccd*B cassette and BsaI restriction sites with an outwards orientation. Flanking this *ccd*B cassette are homology arms and BsmB1 sites. Cloning of multi-transcription units is performed using BsaI, replacing the *ccd*B cassette with transcription unit assemblies released from their respective POT vectors. Digestion of a YAC reveals the homology arms and digestion with BsmB1 releases the five multi-transcription units. Transformation of these DNA fragments into yeast results in its repair by homologous recombination.

In this body of work, I discuss the high-throughput construction of a promoter library, in collaboration with the Dai laboratory from the Shenzhen Institutes for Advanced Technology (SIAT). Here, we endeavour to construct a library of all 6000 promoters from the yeast genome for further characterisation of their activity (Guo et al. 2015). With knowledge of their promoter activities, they can be used for the fine tuning of metabolic pathways for the optimisation of compound production in yeast. The large repertoire of promoters available prevents the repetitive use of the same promoter, decreasing the chances of rearrangements in the pathway from occurring.

3.2 Chapter 3 results

3.2.1 Promoter library construction

This chapter describes the domestication of hundreds of promoters as part of a collaboration between researchers at the University of Edinburgh/Manchester and the Shenzhen Institutes of Advanced Technology towards the publication of the YeastFab toolkit (Guo et al., 2015). As part of the group based in the University of Edinburgh/Manchester, I have domesticated promoters from the *S. cerevisiae* chromosomes II, IV and V into the YeastFab format. As the approach involves high-throughput generation of hundreds of promoters (Figure 3.6), the pipeline was designed to have 4 quality control steps with pass-fail assessments at each phase. In this way, we were able to continually monitor the success rate at the different stages of cloning or validation. This provided the further benefit of facilitating troubleshooting, as an abnormal drop

in success rate at one stage, but not other stages, suggests that the preceding step can be repeated, expediting assembly of larger libraries.

The domestication of all promoters discussed in this chapter (promoters from chromosome II, IV and V) were all carried out by myself. The individual results for the cloning of each promoter in this collectiion was deposited in a digital thesis supplementary folder (JAuxillos_YeastFabChapter3-data).

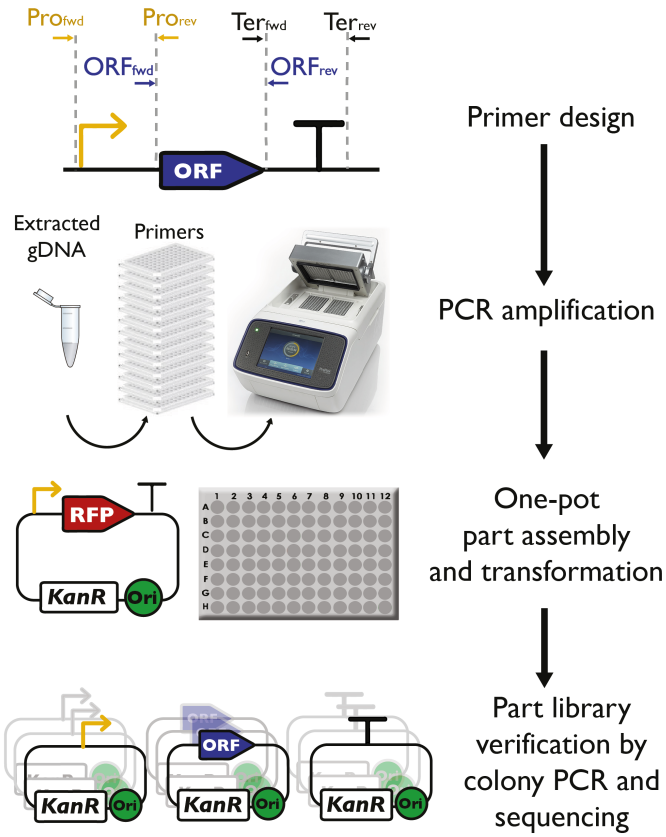


Figure 3.6

The diagram shows the process for the high-throughput domestication of YeastFab parts. Primers are designed to amplify the parts based on the definitions stated in Figure 3.2. These primers are used to amplify the parts from the yeast genome the subsequently domesticated into the YeastFab format by Golden Gate assembly. Parts are verified by colony PCR and Sanger sequencing prior to their storage in the collection.)

In order to keep track of the sample assembly process and the results for each step, the software YeastfabCAM (https://yeastfab.cailab.org) was developed by an informatician, Yisha Luo (Luo et al., 2015). In YeastFabCAM, data for each of the four quality control steps were uploaded and each promoter was scored 'pass' or 'fail' based on a set of criteria for each phase. The quality control steps are 1) PCR, 2) Golden gate assembly and transformation, 3) Colony PCR and 4) Sequencing.

In the PCR step, promoters are curated by comparing the size of the PCR product on the agarose gel to the predicted promoter size. A promoter with a band of the correct size is scored as 'pass'. Amplicons of the wrong size or promoters with no amplicons on the gel are scored as 'fail'

The second quality control step is for the assessment of the golden gate assembly and transformation. The criterion for a 'pass' in this step is whether colonies for each promoter cloning reaction were observed on the plate and if white colonies were present. A 'fail' was scored when no colonies were observed for that promoter cloning reaction or if all colonies were red.

The third quality control step involved the colony PCR verification of 4 clones for each promoter. A 'pass' was scored for promoters where at least 1 amplicon of the correct size was observed on the agarose gel.

The final phase involves the sequencing verification of 1 clone for each promoter. Sequencing trace files are aligned using an in-built feature of the YeastFabCAM database and automatically scoring each promoter as 'pass' based on the following criteria: 1) alignment to the correct promoter in the correct orientation and 2) less than 4 mutations observed for each promoter part. Promoters with sequencing trace files not aligning to the reference sequence, containing more than 4 mutations or with low sequencing quality reads are scored as 'fail'. The arbitrary criterion of less than 4 mutations was based on pre-set design principles defined at the beginning of the collaboration. However, this is not optimal as a mutation in core promoter sequence elements, such as the TATA box or transcription start site (TSS) can affect the promoter activity. Alternatively, mutations within the 5'UTR can alter the sequence composition and secondary structure of the mRNA thereby affecting the regulation of translation by responsive elements.

Table 3.1 outlines the process of construction, showing the percentage success rate at every step of the process, with the colour coding of each cell on the table based on

the percentage success rate (coloured green to red for 100% to 19.4%). The table of quality control results across all 17 plates provides a number of insights into the cloning process. Due to the propensity of the PCR step to produce false negative results, it was decided to proceed in all cases to golden gate cloning followed by transformation into *E. coli* cells. This decision was supported by the evidence that despite there being high variability between the observed success rate for the PCR step between the plates (66.7-99.0%) there was consistently high success rates for the post-transformation RFP screening (95.8-100%). This strongly suggested that the PCR verification did indeed display a high false negative rate and that in future variations upon the strategy this step may be optional. However, the transformation step appeared to disproportionately produce false positive results, with nearly 100% success rate across all 17 plates, as subsequent colony PCR verification shows a drop in the average success rate to 83.3% (Table 1). This bias can be attributed to the criteria for scoring in the transformation step. The criterion for success is based on the presence of white colonies from the transformation of *E. coli* cells with HcKan_P plasmid. This criterion does not however confirm the cloning of the correct promoter. Possible false positive results can be due to the mutation of the RFP cassette in the HcKan_P vector, resulting in a white colony, the re-circularisation of an empty HcKan_P vector or the cloning of an incorrect PCR product, which can all yield in a false positive results. In addition, the accumulation and maturation of RFP to visually identifiable levels can also be a cause of error, with some cases where a discernible reddening of colonies takes longer than 20 hours to appear.

Table 3.1 Percentage success rate at each stage of the promoter library construction pipeline

		Percentage success at the PCR step (%)	Percentage success at the Golden Gate cloning and transformation step (%)	Percentage success at the colony PCR step (%)	Percentage success at the sequencing step (%)
Chromosome 2	Plate 1	89.6	100	84.4	74
	Plate 3	89.6	100	80.2	62.5
	Plate 4	68.8	95.8	87.5	64.6
	Plate 5	79.2	97.2	80.6	48.6
Chromosome 4	Plate 1	97.9	99	69.8	66.7
	Plate 2	99	99	92.7	83.3
	Plate 3	81.3	100	93.8	90.6
	Plate 4	95.8	95.8	93.8	81.3
	Plate 5	89.6	100	91.7	85.4
	Plate 6	97.9	100	89.6	87.5
	Plate 7	91.7	100	89.6	85.4
	Plate 8	96.9	96.9	76	66.7
	Plate 9	94.1	100	98.5	64.7
Chromosome 5	Plate 1	68.8	100	70.8	42.7
	Plate 2	71.9	100	54.2	35.4
	Plate 3	93.8	99	83.3	29.2
	Plate 4	66.7	100	80.6	19.4

Promoters not successfully amplified in the PCR step (Fail result in the PCR step) should not result in any successful results in downstream steps. In this analysiss, I aim to validate this result. Figure 3.7 is an analysis of the Fail score for two quality control steps (Step 1 - PCR and Step 3 - colony PCR verification) in two exemplar promoter cloning platess. Although the success rates described in Table 3.1 of the PCR and colony PCR step appear similar, further analysis of the failed promoters in both steps in two exemplar plates show a low correlation, with only a handful that fail in both cases (black dots) (Figure 3.7).

The high number of promoters which were scored a 'fail' in the PCR step (step 1) but showed a 'pass' result in the later step (Step 3 - colony PCR) refutes the initial hypothesis that promoters which have not shown to pass in the PCR step should not result in any successful results in downstream steps. In particular, the high incidence of promoters which fail in the PCR step (labelled in blue in Figure 3.7) suggests the propensity of the PCR verification step to produce false negative results.

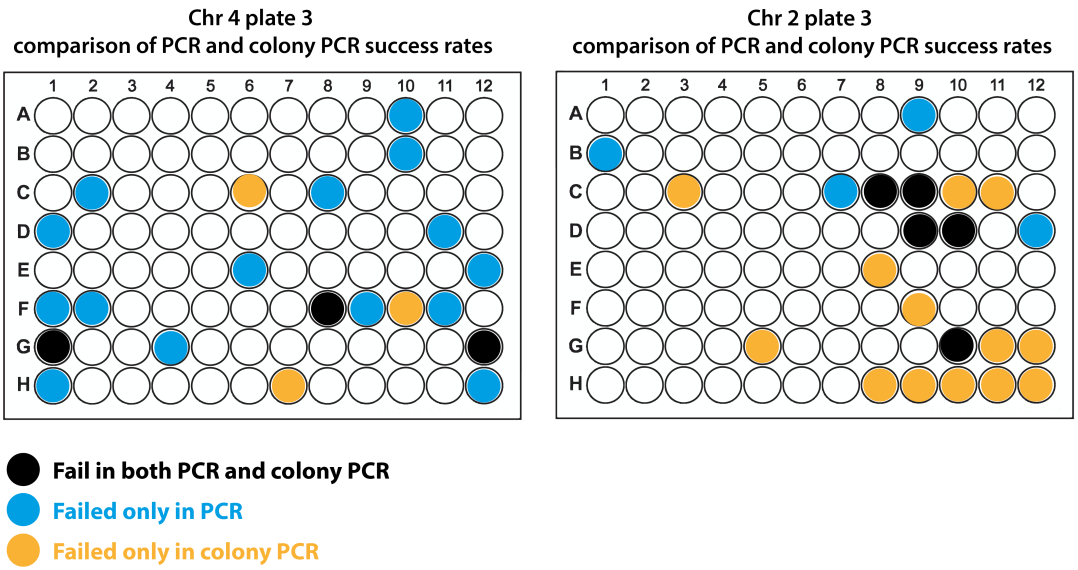


Figure 3.7

Comparison of promoters that have an annotated fail score in the PCR and colony PCR steps in in Chr 4 plate 3 and Chr 2 plate 3 (analysis was based on data shown in Appendix Figure C.1 and C.2). This analysis revealed promoters which have been annotated with a Fail during both PCR and colony PCR steps are the promoters which have be

Across all 4 steps, the variability of the sequencing success rate in all 17 plates in sequencing appears the highest, with the lowest percentage of 20 and the highest reaching 90 (Figure 3.8). This could be due to a more precise method of validation of cloning successes, allowing the verification of multiple factors such as the alignment of the correct promoter, correct cloning orientation and less than 4 mutations. Several factors may have also contributed to the high failure rate, such as poor sequencing quality of the entire promoter, low sequencing quality at the beginning of the read, incorrect base annotation or presence of multiple peaks in the sequencing chromatogram. All of these however result in the reporting of a false negative result. A limitation of the automatic analysis of sequencing results is that it does not differentiate between a true negative result (where a cloned promoter has a high number of mutations or the incorrect promoter assembled into the HcKan_P vector) and a technical issue with the sequencing step therefore only requiring the resequencing of a sample. Therefore, an additional manual verification was required for the curation of sequencing results.

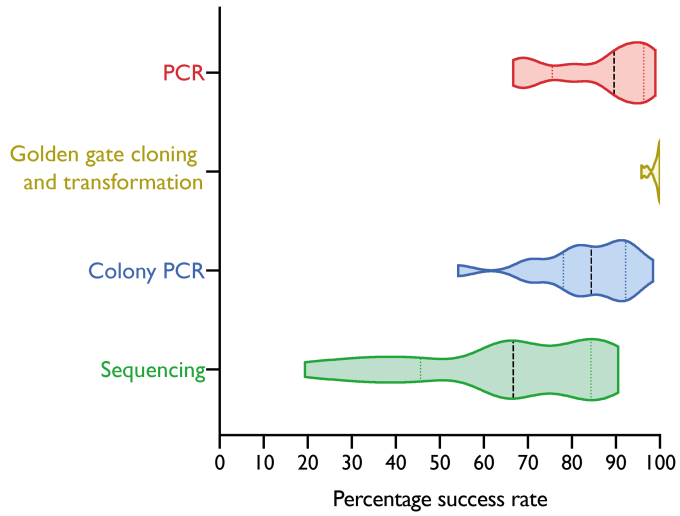


Figure 3.8

The distribution of success rates of each step across all 17 promoter construction plates

Another feature that can possibly affect the efficiency of cloning is the presence of the BsaI restriction site in the promoter sequence. As the promoters were not synthesised and were instead amplified from the yeast genome, the restriction sites could not be excluded. As an example, the chromosome IV collection of 836 promoters contains 37 with an internal BsaI restriction site. Among these 37, more than half were

successfully cloned (20 out of 37 - 54.1% success rate). However, this is significantly lower than the average promoter cloning success rate (79.5%) in the chromosome IV collection (Figure 3.9). Further investigation into the success rate at each step of the cloning process showed a drop during the colony PCR step, with several promoters producing an amplicon band of 200 bp corresponding to the circularization of an empty HcKan_P vector. Although it is still possible to clone promoters with internal BsaI sites into the HcKan_P vector, the efficiency is lower therefore requiring the screening of more clones to be able to isolate a clone with the correct assembly. Optimisations described later in the chapter can seek to better filter out false positive results at the second quality control step (golden gate assembly and transformation) and allowing for the screening of more potentially correct assemblies.

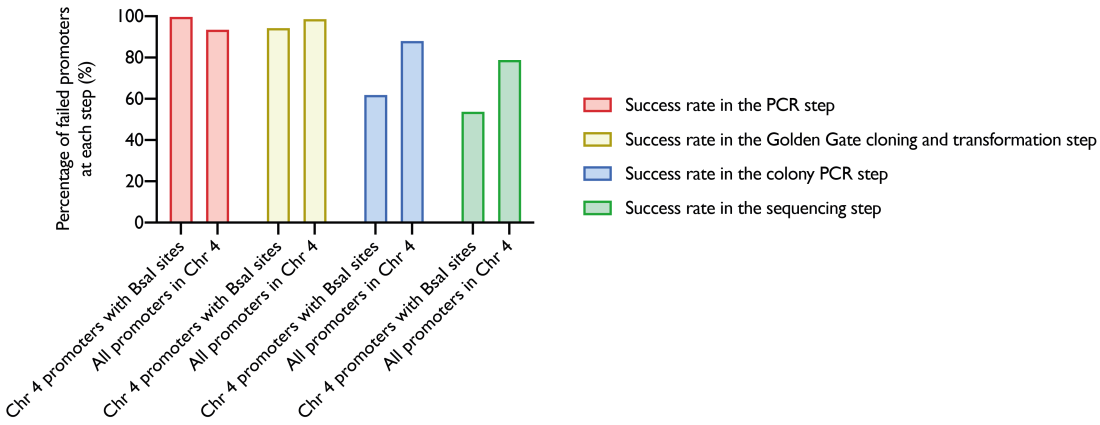


Figure 3.9

Identifying the effects on internal BsaI sites. A) comparison of promoter construction success rates at each quality control step between promoters containing BsaI sites and all promoters in Chr 4.

3.2.2 Modification of the HcKan vectors to allow *ccd*B-based screening

The HcKan_P, HcKan_O and HcKan_T parts accepting vectors constructed as part of the YeastFab toolkit contained an RFP cassette flanked by BsmBI restriction sites with specific 4 bp overhangs. This RFP cassette served as the insertion site for a promoter, ORF and terminator, using red-white screening for isolating clones with the correctly assembled vector. However, this often resulted in false positive results due to the possible mutation of the RFP cassette in the HcKan_P vector, resulting in a white

colony or the re-circularisation of an empty HcKan_P. The reddening of colonies from the accumulation and maturation of RFP would often take longer than 20 hours to appear causing an error in reporting (Balleza et al., 2018). Therefore, other methods of screening such as the *ccd*B-mediated positive selection is preferable to improve the efficiency of the cloning process (Figure 3.10A)

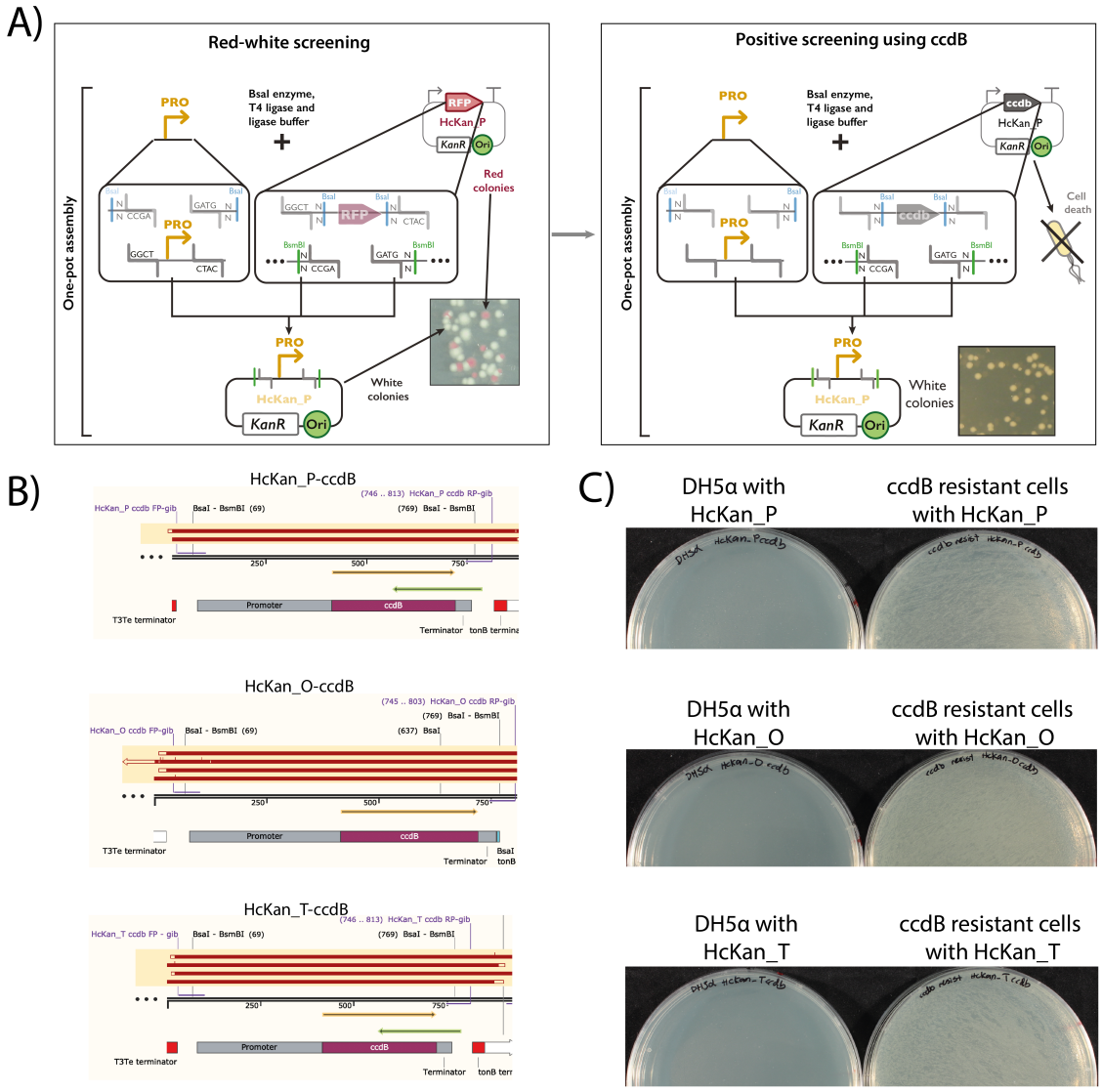


Figure 3.10

A) Modification of the screening method from red-white based to a positive selection using *ccd*B. B) Sequencing verification of the all three parts accepting vector (HcKan_P, HcKan_O and HcKan_T), replacing the RFP with a *ccd*B cassette. C) Verification of the activity of *ccd*B as a toxic gene via transformation into DH5αcells.

In the following work, I have modified all parts acceptor vectors (HcKan_P, HcKan_O and HcKan_T) were modified to replace the RFP cassette with a *ccd*B cassette (Figure 3.10B). This enables positive selection as *ccd*B is a toxic gene that produces a protein product which acts on DNA gyrase when in complex with DNA, providing a barrier to transcription by RNA polymerase (Critchlow et al., 1997; Dao-Thi et al., 2005) *E. coli* cells (such as TOP10, *E. coli* 10 β , Mach1, DH5 α cells) that contain a plasmid harbouring *ccd*B are not able to survive. Therefore *ccd*B-containing vectors require a special host strain for their propagation (such as *ccd*B Survival™ 2 – Table of bacteria strains).

In order to replace the RFP cassette with the *ccd*B cassette for all three parts acceptor vectors, three pairs of primers (HcKan_P *ccd*b FP-gib and HcKan_P *ccd*b RP-gib for HcKan_P, HcKan_O *ccd*b FP-gib and HcKan_O *ccd*b RP-gib for HcKan_O and HcKan_T *ccd*b FP-gib and HcKan_T *ccd*b RP-gib for HcKan_T) were designed to amplify the *ccd*B cassette from the Re2.8 reporter plasmid (Table of primers). These primers were designed to incorporate the unique prefix and suffix sequences for the three respective parts acceptor vectors, alongside 40 bp homology arms to enable construction by Gibson assembly. Sequencing verification of the HcKan_P, HcKan_O and HcKan_T vectors using the HcKan chc FP and RP primers showed correct assembly of the cassette with no mutation in the *ccd*B gene (Figure 3.10B). Verification of the functionality of the *ccd*B gene was carried out by transforming the vectors into both DH5α cells and *ccd*B resistant *E. coli* cells (*ccd*B Survival 2 - Table of bacteria strains). The presence of kanamycin in the lysogeny broth (LB) selects for DH5α cells containing the HcKan_P-*ccd*B plasmid, however the presence of the *ccd*B cassette results in transcriptional inhibition. *ccd*B resistant *E. coli* cells harbouring the HcKan-*ccd*B vector are the only ones that can survive and propagate on LB agar containing kanamycin. Figure 3.10C shows no growth of DH5α cells when transformed with the HcKan-*ccd*B vector and selected on LB agar containing kanamycin, for all 3 vectors tested (Figure 3.10C). Here, I have shown the modification the YeastFab assembly method through the construction oof parts-accepting vector containing the *ccdb* cassette instead of the RFP cassette, allowing the use of positive selection system rrather than red-white screening.

3.2.3 Modification of the POT vectors to allow *ccd*B-based screening

The POT1-RFP vector in the YeastFab toolkit for transcription unit assembly also contains an RFP cassette. This is flanked by BsmBI restriction sites with specific 4 bp overhangs (GGCT as the prefix and CCTC as the suffix), which serves as the insertion site for a transcription unit (promoter-ORF-terminator). However, similar problems were encountered with the use of red-white screening for golden gate assemblies. As a result, the POT vector was modified to replace the RFP cassette with a *ccd*B cassette (Figure 3.11A). Primers were designed for the amplification of the *ccd*B cassette from the HcKan_P vector with 40 bp homology to the POT backbone vector and cloned by Gibson assembly (POT1-*ccd*B (-BsaI) Gibson FP and POT1-*ccd*B (-BsaI) Gibson RP). The resulting plasmid was sequence verified using M13 primers and showed the correct assembly of the *ccd*B cassette into the POT vector (Figure 3.11B).

The same replacement was carried out for 8 POT vectors (POT2-RFP, POT3-RFP, POT4-RFP, POT5-RFP, POT6-RFP, POT7-RFP, POT8-RFP and POT9-RFP) in the YeastFab toolkit. A different set of primers were ordered for each POT vector due to the presence of a different set of prefix and suffix flanking the cloning site (Table of primers). These assemblies were sequence verified, showing the correct assembly with no mutations introduced (Figure 11B).

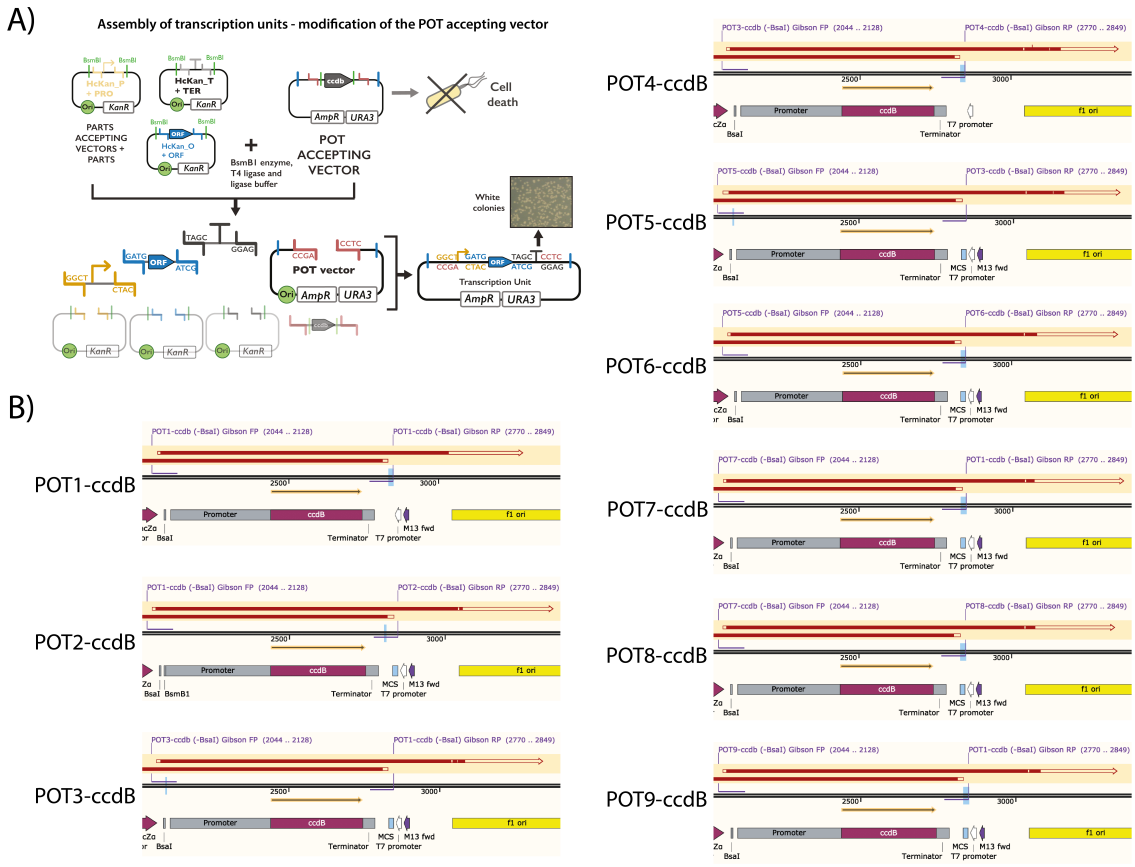


Figure 3.11

A) Modification of the screening method from red-white based to a positive selection using *ccd*B. B) Sequencing verification of the all POT acceptor vector (POT1, POT2, POT3, POT4, POT5, POT6, POT7, POT8 and POT9), replacing the RFP with a *ccd*B cassette

3.3 Chapter 3 discussion

This chapter describes work carried out in collaboration with the Dai laboratory towards the publication of the YeastFab toolkit. Here, I have constructed a library of promoters from chromosome II, IV and V resulting in the domestication of hundreds of promoters for further activity characterisation (Chapter 4) and utilisation for the expression of biosynthetic pathway genes in yeast (Chapter 5 and 6).

The analysis of success rates in each step shows that the initial PCR analysis does not appear to be essential as it appears to show a high rate of false negative results.

Therefore, unlike the other quality control steps, all promoters were taken forward to the next cloning step irrespective of the PCR results. In contrast, the golden gate and transformation quality control step appears to show false positive results as the criterion for success relies on presence of white colonies from the transformation of *E. coli* cells with HcKan_P plasmid. This criterion, however, is not dependent of the cloning of the correct promoter. Possible false positive results can be due to the mutation of the RFP cassette in the HcKan_P vector, resulting in a white colony or the re-circularisation of an empty HcKan_P vector which can also result in a false positive result. In addition, the accumulation and maturation of RFP to visually identifiable levels can also be a cause of error, with some cases where a discernible reddening of colonies takes longer than 20 hours to appear. A possible improvement for the protocol can involve the incorporation of a digestion step at the end of the cycling Golden Gate program to linearise all plasmids containing internal BsaI or BsmBI sites followed by the specific removal of linear double-stranded DNA using PlasmidSafe ATP dependent DNase (Epicentre).

Further verification of the assemblies by colony PCR and sequencing show a decline in success rates. One factor was investigated to analyse its effect on success rates – the presence of internal BsaI sites. Results from this analysis show a slight decrease in cloning efficiency, however correct assemblies were still obtained. Future work can involve the removal of genetic parts containing internal BsaI or BsmBI sites. Alternatively, alternative options to improve success rates can be carried out such as the optimisation of the Golden Gate assembly cycling program or the screening of a higher number of clones to obtain one with the correct assembly.

The criteria for a successfully assembled promoter consist of its alignment to the reference sequence, correct cloning orientation and less than 4 mutations. The decision for tolerating 4 mutations is due to two reasons, 1) as these promoters are amplified from the yeast genome, it is plausible that mutations existed in the strain and 2) mutations can introduced during the PCR amplification step. Improvements in the PCR step, such as the use of an improved polymerase with higher fidelity can reduce the expected mutation rate. However, as previously mentioned, the presence of these mutations on elements with high sequence conservation can result in deleterious effects on the promoter activity. In particular, changes in the secondary structure of the mRNA, in particular in the 5'UTR, can result in different protein output. Dvir et al. showed that 5'UTR variants that form a stable secondary structure hamper translation initiation, resulting in a lower protein production (Dvir et al., 2013). They also show

that presence of out-of-frame upstream AUGs (uAUGs) affect translation efficiency, possibly due to the presence of premature stop codons (Dvir et al., 2013). Future work involving the expansion of the promoter library can involve the construction of promoter variants without the mutation and comparing the expression profile between mutated and wild type variants.

In the second half of this chapter, I describe efforts to modify the YeastFab toolkit, replacing the red-white screening method following the golden gate assembly, to a *ccd*B-mediated positive selection system. All parts acceptor vectors (HcKan_P, HcKan_O and HcKan_T) along with POT vectors (POT 1-9) were modified to replace the RFP with a *ccd*B cassette. Therefore, in this system, common cloning *E. coli* cells (such as TOP10, *E. coli* 10β, Mach1, DH5α cells) harbouring a plasmid containing the *ccd*B toxic gene are unable to survive. Cells that are able to propagate in LB agar with kanamycin selection are those that harbour a yeast promoter cloned into the HcKan_P plasmid. Further verification will need to be carried out to identify the efficiency of cloning by screening all clones on a plate and identifying the percentage of colonies containing the correct assembly.

Chapter 4

High-throughput characterisation of promoters

4.1 Chapter 4 introduction

4.1.1 Methods for optimising bioproduction

When engaging with metabolic engineering, a number of factors must be considered when seeking to generate a desired compound and optimising production. Whilst the selection of the appropriate enzyme coding sequences is essential, the expression level modulation of each enzyme within the pathway also plays a central role in the capacity to avoid high toxicity, achieve detectable levels of the end compound and ultimately optimise the titres of the end compound.

For tuning gene expression within pathways there are a number of potential approaches that can be considered in yeast. The first set of approaches centre around nascent transcriptional control. These approaches include alternative promoter selection alongside the use of inducible 'On' systems to enable over expression or the use of inducible 'Off' systems to enable repression. When considering On or Off systems we can use more traditional approaches derived from endogenous transcription factors such as *LacI* or more recent approaches relying upon programmable DNA binding domains such as zinc finger (McIsaac et al., 2014, 2011) and dCas9 derived synthetic transcription factors or repressors (Zalatan et al., 2015). Alternatively, post-transcriptional

control can be achieved using ligand sensitive RNA based riboswitches, which induce conformational changes within the RNA in the presence of the relevant ligand. This can be used to expose, or block key RNA elements involved in translation initiation. Finally post translational control can be achieved by fusing inducible degron tags such as the auxin inducible degron to an enzyme of interest to enable inducible degradation (Apel et al., 2017; Mendoza-Ochoa et al., 2019).

4.1.2 Regulation of gene expression

It is important to note when considering the following body of work that much of the efforts towards the generation of standardised parts libraries have been based upon a principle of attempting to discretise continuous stretches of DNA that have co-evolved over many millennia. As such isolated promoter parts can be expected to lack some *cis*-regulatory elements due to the sequence length range used by each publication in defining a promoter sequence. However, for applications such as metabolic engineering they provide a very useful tool for variable expression profiles for larger pathways.

Whilst a clear-cut definition of a 'promoter' proves challenging due to complicated roles played by variably conserved proximal and distal genetic elements, here we define a promoter as a stretch of DNA that can trigger and regulate the initiation of transcription (Ottoz and Rudolf, 2018). Yeast promoters can be fragmented into two distinct functional regions: the core promoters and upstream regulatory elements (Lubliner et al., 2015; Struhl, 1982). The core promoter region contains the essential elements required to initiate transcription, such as the canonical TATA box or TATA-like sequences and the transcription initiation start site (TSS)(Blazeck and Alper, 2013) (Figure 4.1). A suggested consensus TSS sequence are $A(A_{rich})_5NPyA(A/T)NN(A_{rich})_6$ (with A representing the first transcribed base) and RRYRR,TCRA,YAWR (Lubliner et al., 2015; Zhang and Dietrich, 2005). The presence of multiple TSS in the core promoter can result in the transcription of multiple unique isoforms of the same gene (Pelechano et al., 2013). Upstream regulatory elements can modulate promoter activity through the recruitment of transcription factors. Upstream activating sequences (UAS), which stimulate transcription initiation, contain transcription factor binding sites (TFBS) (Brindle et al., 1990; Carmen et al., 1995; Guarente et al., 1984). For many promoters commonly used for metabolic or pathway engineering, this includes the 5'UTR along with the core promoter region

The yeast promoters currently utilised for molecular biology can be subdivided into two classes, constitutive promoters and regulated promoters.

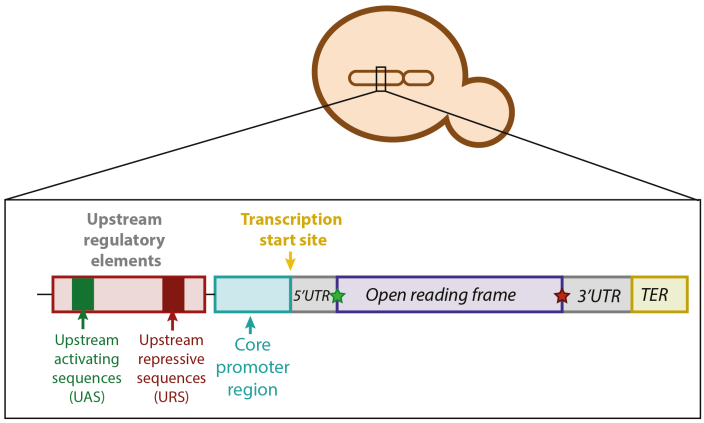


Figure 4.1

A diagrammatic representation of the structure of a yeast promoter. The core promoter is the main determinant of gene expression in yeast, containing essential elements such as the TATA box (or TATA-like sequences) (Blazeck and Alper, 2013). This serves as the location for the formation of the pre-initiation complex and subsequent scanning for an appropriate transcription start site (TSS). Upstream of the core promoter region is a segment of DNA containing upstream regulatory elements which can modulate promoter activity through the recruitment of transcription factors to upstream activating sequences (UAS) or upstream repressive sequences (URS) (Brindle et al., 1990; Carmen et al., 1995; Guarente et al., 1984). The 5'UTR is located downstream of the core promoter region. For many promoters commonly used for metabolic or pathway engineering, this includes the 5'UTR along with the core promoter region

4.1.3 Constitutive promoters

As described by Ottoz and colleagues, constitutive promoters are those that possess a relatively constant expression profile over time (Ottoz and Rudolf, 2018). Overexpression of proteins in yeast has centred around a few well characterised and highly expressed promoters. Many of these constitutive promoters are taken from genes involved in the yeast glycolytic pathway (pPGK1 - for phosphoglycerate kinase, pPDC1 – pyruvate decarboxylase, pTPI1 – triosephosphate isomerase, pADH1 – alcohol dehydrogenase 1, pTDH3 – glyceraldehyde-3-phosphate dehydrogenase). In particular, pPGK1 and pTDH3 are the strongest known promoters and are known to be regulated

by the same set of transcription factors, Rap1 and Gcr1 (Ottoz and Rudolf, 2018). Other widely used promoters are pTEF1 (translation elongation factor), pACT1 (β-actin), pHXT7 (hexose transporter), pADH1 and pCYC1 (Da Silva and Srikrishnan, 2012; Mumberg et al., 1995; Ottoz and Rudolf, 2018). The promoter pCYC1, driving the expression of cytochrome c isoform 1 has been robustly characterised. pCYC1 is naturally repressed in the presence of glucose, mediated by UAS1 and UAS2. However Guarente *et al.* showed that mutation of UAS2 resulted in an increase in the promoter activity by 10-20 fold in the presence of glucose (Guarente et al., 1984). Similarly, the 1500 bp promoter sequence of pADH1 was shown to be active in the presence of glucose and in exponential phase and inactive upon the shift to ethanol utilisation (Denis et al., 1983; Ruohonen et al., 1991). Ruohonen *et al.* then identified a modification in the expression profile of the pADH1 promoter following a truncation of 1100 bp, resulting in a delayed upregulation (towards the early stages of ethanol utilisation) (Ruohonen et al., 1991).

Whilst these handful of promoters have proven effective for the overexpression of a single protein, a larger repertoire of promoters is required for effectively coordinating the expression of biological circuits and more complicated multi-gene metabolic pathways. In addition, incorporating variation in the promoter sequences used for multi-gene assemblies can reduce the risk of unwanted homologous recombination resulting from repetitions of the same promoter sequence. A wider range of promoter activities is also required to enable fine tuning of gene expression to the appropriate levels to optimise increasingly complex genetic programs.

More recent advances in the field of promoter engineering have attempted to address these requirements. Several examples include the alteration of the promoter transcription regulation capabilities by mutating the sequence of commonly used promoters resulting in the generation of promoter libraries possessing ranges of expression. Alper and colleagues utilised error-prone PCR to generate a library of mutant *TEF1* promoter sequences, going on to show a diversity of expression in exponential phase, with many mutants showing reduced activity, but several also displaying increased (up to $\sim$3 fold higher) expression.

Alternatively, researchers have explored the possibility of utilising a larger subset of promoters derived from the yeast genome. Native yeast promoters orchestrate a wide range of expression programs, impacting not only the average expression of the associated coding sequences but also helping drive the relative expression at different

stages of culturing. Lee and colleagues isolated 19 native yeast promoters (which they defined as the 700 bp sequence upstream of their corresponding native gene) from the yeast genome as part of the Yeast MoClo kit and characterised their activity, measuring fluorescence at a single time-point. These promoters were shown to span a wide range of strengths, with a ∼500-fold(Lee et al., 2015). Lee and colleagues found pTDH3 ranking as the most active promoter across all 19 screened and also observed relatively high promoter activity with pPGK1, pTEF1 and pTEF2, confirming previous findings by Partow et al (Partow et al., 2010). Similarly, Apel and colleagues (2017) screened 37 native yeast promoters (which they defined as the 600 bp upstream of the ATG of the corresponding native ORF). These 27 promoters are from genes involved in multiple yeast metabolic pathways, including glycolysis, the TCA cycle and stress response. They utilised flow cytometry to characterise promoter activities at four time points (4, 8, 24 and 48 hours) in different yeast growth media (YPG, YPD and CSM). Promoters screened over the different time points and media conditions displayed an ∼100-fold range in fluorescence (Apel et al., 2017). Similarly to Lee *et al.*, Apel and colleagues saw some of the highest expression from the TDH3 promoter, however they saw significantly reduced fluorescence at 24 hours compared to 4 hours (34% reduction in fluorescence) (Apel et al., 2017). Apel and colleagues observed that the activity of pPGK1 appeared to maintain the same level of expression across multiple time points and growth media. They also determined that promoters commonly used in protein overexpression or pathway engineering endeavours (pTEF1 from a translation elongation factor and pENO2 from a glycolytic enzyme) appeared to show high exponential phase specific activity, with a rapid decline in activity towards the stationary phase. Several stress-response associated promoters were also identified to possess elevated expression in stationary phase (pHSP26, pHSP104 and pSSA1).

Guo *et al.* sought to explore a much larger range of native promoter sequences, analysing hundreds of promoters (defined as the 500 bp sequence upstream of the ATG). Currently this library is one of the most comprehensive collections of yeast promoter parts available for the assembly of pathways and genetic circuits. Guo *et al.* characterised the activities of all promoters in exponential phase (OD = 1), grown in normal conditions (growth in 30°C in Sc-Leu media) and stress conditions (nutrient deprivation, heat or oxidative stress) (Guo et al., 2015) (Figure 4.2). They found a large representation of weak promoters in the collection. Stress was clearly shown to induce differences in promoter activity, with corresponding changes in protein abundance. Oxidative stress appeared to result in the upregulation of over half of the promoters screened while heat stress led to the downregulation of most promoters

tested. Interestingly, when comparing glucose and nitrogen stress, similar trends in promoter activities were observed. Further validation and characterisation is required to identify the cause of the similarity in promoter activities across these two conditions.

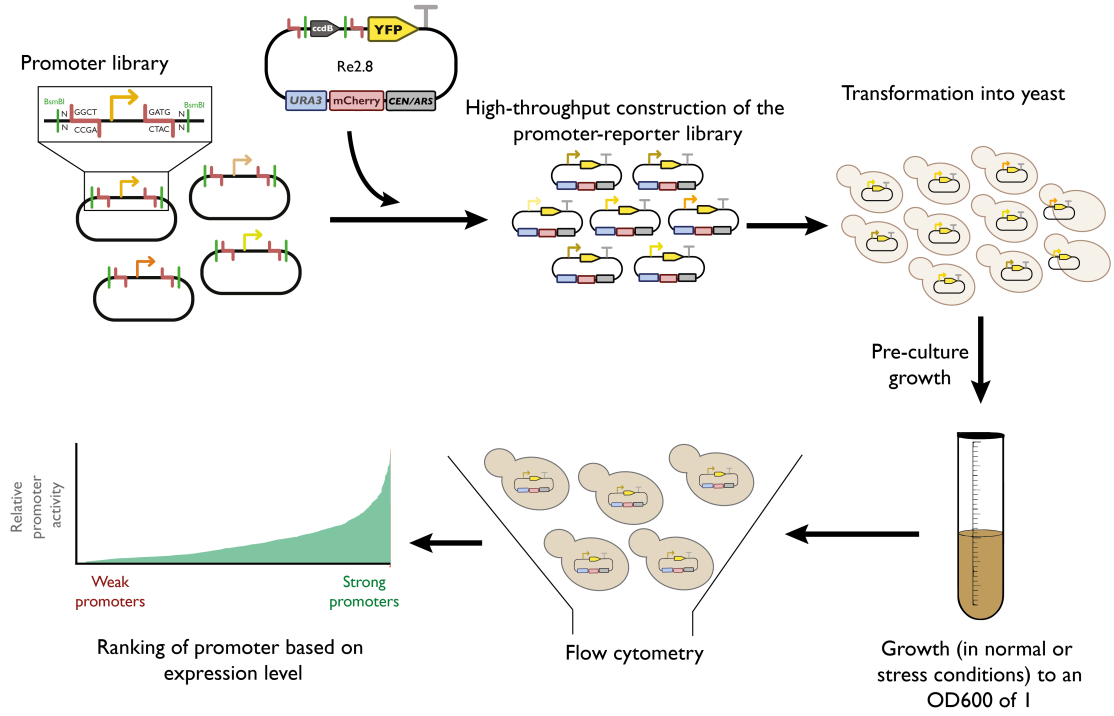


Figure 4.2

The promoter construction and characterisation pipeline used by Guo *et al* (2015). The pipeline begins with the high throughput cloning of yeast promoters into the Re2.8 a dual fluorescence reporter vector. Successfully assembled promoter-reporter plasmids are transformed into yeast and selected in Sc -Leu media. These yeast cells containing the reporter vectors are grown in normal (SC-LEU media) or stress conditions to exponential phase (OD600 of 1) and analysed by flow cytometry. The relative expression was calculated by normalising the YFP by the mCherry fluorescence measurements. In order to rank the promoter based on activity, the normalised YFP fluorescence of each promoter was divided by the normalised YFP fluorescence of pCYC1.

4.1.4 Regulatable promoters

Whilst such libraries of promoters enable some measure of control over the strength of expression for genes within a pathway, in some cases it will be desirable to achieve temporal control of expression; for example, to coordinate the expression of a toxic

enzyme based upon the availability of substrate(s). This becomes possible when utilising inducible systems, which can be applied to prevent the accumulation of a toxic enzyme or enable increased expression of a downstream enzyme after sufficient precursors have accumulated.

Inducible promoters refer to those where the expression levels can be controlled by the presence of a small molecule or synthetic transcription factor. Desirable features sought out in these inducible systems include; 1) no activation of the target gene without the presence of the inducer or expression 'leaking', 2) rapid upregulation of target gene expression upon induction with the inducer and 3) the absence of growth defects or unwanted modification of the cell's metabolism upon induction.

In yeast, the most commonly used inducible promoters belong to the GAL system. The Gal4 protein functions as the transcriptional activator. This protein consists of a DNA binding domain that recognises an UAS, a glucose response domain (GRD) and an activation domain (Stone and Sadowski, 1993) (Figure 4.3A). Gal80 acts an inhibitor of Gal4, binding in the absence of galactose thereby preventing the activation of gene expression. However, upon the introduction of galactose, Gal80, in complex with Gal3 disassociates from Gal4 allowing the binding of Gal4 to the UAS sequence within the promoter region of *GAL1*, *GAL7* and *GAL10* resulting in transcriptional initiation (Figure 4.3B) (Stone and Sadowski, 1993). Transcript levels of *GAL1* and *GAL10* increase by ∼1000-fold upon the induction with galactose compared to the expression levels without galactose present (Adams, 1972; St John and Davis, 1981). However, in the presence of glucose, as and as it is the preferred carbon source, transcriptional activation of *GAL1* is inhibited by GRD and Mig1 (Rohde et al., 2000; Stone and Sadowski, 1993). The Gal4-mediated inducible promoters have been widely used in both the overexpression of genes of interest and the temporal induction of biosynthetic enzyme genes for the production of compounds in yeast. To utilise this system, genes of interest are cloned downstream of promoters (ie. pGAL1, pGAL10) containing the UASGAL CGG-N11-CCG motif (Giniger et al., 1985), which is recognised by the homodimer of the Gal4 protein (Marmorstein *et al.*, 1992; Hong *et al.*, 2008). Presence of glucose in the Gal4 results in transcriptional inhibition mediated by the GRD within the Gal4 protein (Rohde et al., 2000; Stone and Sadowski, 1993) (Figure 4.3C).

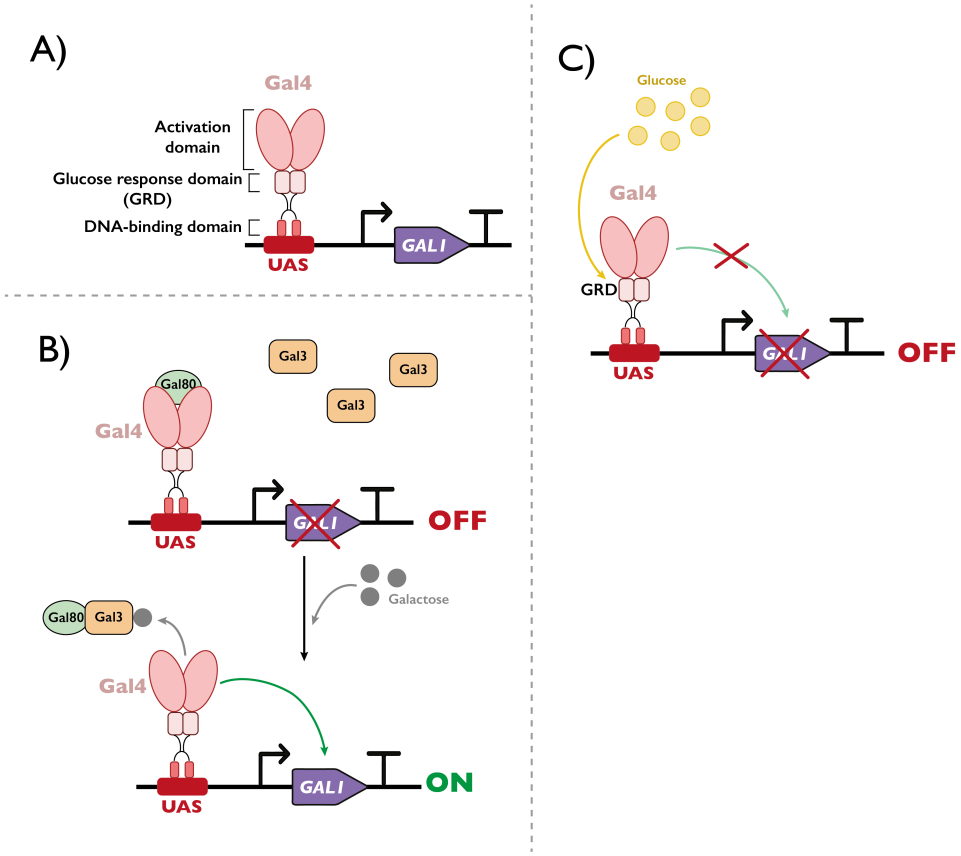


Figure 4.3

A diagrammatic representation of the GAL system. A) The Gal4 protein can be subdivided into two main regions, the activation and the DNA binding domain. B) This Gal4 transcriptional activator recognises the upstream activating sequence (UAS) CGG-N11-CCG motif upstream of the GAL promoters. Several targets for the Gal4 activator are present throughout the yeast genome. C) Without galactose, the inhibitor Gal80 associates with Gal4, preventing transcriptional activation. Introduction of galactose results in the disassociation of Gal80, in complex with Gal3, allowing transcriptional initiation. Introduced glucose interacts with the GRD domain inhibiting transactivation by Gal4 of the downstream gene.

4.1.5 Riboswitch-based regulation of protein production

Protein production can also be controlled post-transcriptionally, by engineering mRNA to be translationally inactive in the presence of a small molecule. In the past decade, there has been an increasing understanding of the complexity of RNA-mediated regulation of cellular processes. The discovery of different classes of regulatory

RNAs (such as miRNA, small and long ncRNAs) has been shown to orchestrate various processes from translation regulation, and adaptation to stress conditions to defence against viral infection. Another class of regulatory RNA are riboswitches, commonly found in prokaryotic organisms. Segments of RNA within the 5'UTR of an mRNA function as a molecular switch that regulates transcription, mRNA processing or translation (Groher and Suess, 2014). Modulation of cellular processes by the mRNA occurs through conformational change of the RNA structure, mediated by the interaction with a specific small molecule. The understanding of these principles for riboswitch-mediated regulation has prompted researchers to engineer and optimise synthetic riboswitches for controlling gene expression in different chassis such as *S. cerevisiae*.

Several riboswitches have been developed for the regulation of translation in eukaryotes. Engineered riboswitches rely on the action of aptamers, short single stranded oligos that can interact with specific molecules, for sensing and undergoing a conformational change in the presence of small molecules. Incorporation of these aptamers to the 5'UTR sequence of a reporter gene leads to the reversible repression of protein production in the presence of the small molecule, in a dose-dependent manner. A major advantage is that aptamers can be screened for sensitivity to any desired target small molecule. This screening is achieved through the systematic evolution of ligands by an exponential enrichment (SELEX) method (Ellington and Szostak, 1990; Tuerk and Gold, 1990). However, many aptamers screened *in vitro* display no regulatory activity upon *in vivo* testing. To date, only a handful of aptamers have been incorporated as part of a functional riboswitch for the regulation of gene expression (Groher and Suess, 2014). Analysis of a pool of neomycin aptamers by NMR spectroscopy showed that high binding of the aptamer to the ligand was not the only pre-requisite for regulatory function in a riboswitch context. They found that variants of the neomycin aptamer that failed to regulate gene expression showed no conformational changes in the presence of the ligand (Weigand *et al.*, 2011). Groher *et al.* noted that these failed variants dominated the population of enriched aptamer pools in the in vitro selection by SELEX suggesting the crucial requirement of an in vivo screening protocol as part of riboswitch engineering (Boussebayle et al., 2019; Groher and Suess, 2014).

The position of the aptamer in the 5'UTR has been shown to significantly impact the efficiency of riboswitch regulation. Aptamers inserted immediately downstream of the cap structure result in a conformational change in the presence of the small molecule,

inhibiting the binding of the small ribosomal subunit to the cap structure (Hanson *et al.*, 2003). In contrast, insertion of the aptamer distal to the cap structure and upstream of the AUG start codon blocked the scanning process due to the formation of a secondary structure by the aptamer in the presence of the target molecule, ultimately preventing 80S ribosome formation (Suess *et al.*, 2003). Between these two, Hanson et al showed that cap proximal aptamers were significantly less active than cap distal aptamers (Hanson et al., 2003). Mutation of different sequences within the aptamer has also been shown to improve its regulatory properties (Weigand et al., 2008). The activity of these aptamers has been tested in the context of pTDH3 and pADH1, showing variable regulatory properties and transcriptional activation strength mediated by the promoter (Kötter *et al.*, 2009)

Chapter 3 described work previously carried out to establish the YeastFab assembly toolkit, including the domestication of several hundred promoters. In previous publications, the activity of selected promoters were only assessed by measuring the activity at a single time point chosen to represent exponential phase, with the exception of Apel *et al.*, who demonstrated the variability in expression that can be observed from measuring 4 time points (Apel et al., 2017). Guo *et al.* provided a very large library for subsequent use for pathway engineering, however they did not explore the dynamics of promoter expression observed over the course of culturing (Guo et al., 2015). In the following work I build upon their characterisation, by exploring a subset of their library, analysing expression profiles temporally. This is motivated by the context within which metabolic engineering occurs, with compounds of interest being produced over several hours of culturing during which cells would be expected to transition from exponential phase through to stationary phase. After initial characterisation of a subset of highly used promoters, I apply the established screening pipeline to characterise 74 promoters within the YeastFab library, previously analysed only at late exponential phase, exploring their activity profile over time. In subsequent work, I characterise two regulatable systems for the purposes of obtaining temporal control of pathway expression. I explore a synthetic transcription factor based 'On' system and a neomycin riboswitch based 'Off' system, which is subsequently applied for pathway expression in chapter 6.

4.2 Chapter 4 results

4.2.1 Characterisation of promoter expression profiles over time

Promoters from the collection previously characterised by (Apel et al., 2017), along with common yeast promoters previously constructed by Dr. Wei Liu from the Cai group (unpublished data) were characterised further using a microplate reader. The annotated function for the corresponding gene of each promoter is outlined in Table 4.1, with several promoters in this pilot screen belonging to ribosomal proteins (RPS3, RPS13 and RPS2) and the glycolysis pathway (TDH3, PGK1).

Table 4.1 Promoter information and key functions of their corresponding ORFs (obtained from SGD - https://www.yeastgenome.org/

Promoter name	Promoter ID	ORF name aliases	Key description of ORF
pTDH3	YGR192C_P	TDH3,GLD1, HSP35, HSP36,SSS2	involved in glycolysis and gluconeogenesis
pPGK1	YCR012W_P	PGK1	key enzyme in glycolysis and gluconeogenesis
pACT1	YFL039C_P	ACT1,ABY1, END7	Actin, structural protein involved in cell polarisation
pRPS2	YGL123W_P	RPS2,RPS4, SUP138, SUP38,SUP44	Protein component of the small (40S) subunit,
pZEO1	YOL109W_P	ZEO1	Peripheral membrane protein of the plasma membrane and is involved in regulating cell integrity pathway
pCYC1	YJR048W_P	CYC1	Electron carrier of the mitochondrial inter membrane space
pRPS3	YNL178W_P	RPS3,SUF14	Protein component of the small (40S) ribosomal subunit
pRPS13	YDR064W_P	RPS13,RPS13B, RPS13C	Protein component of the small (40S) ribosomal subunit
pHSP26	YBR072W_P	HSP26	Small heat shock protein (sHSP) with chaperone activity

Six (pTDH3, pPGK1, pACT1, pRPS2 and pZEO1, pCYC1) of these promoters were previously domesticated to the YeastFab standard. The remaining three (pRPS3, pRPS13 and pHSP26) were amplified from the yeast genome (Table of primers) and

cloned into the HcKan_P vector by Golden Gate assembly. All promoter parts were verified by Sanger sequencing, showing both successful cloning into the correct vector and no introduction of mutations into the promoter sequences. These promoters were subsequently cloned into the Re2.8 dual fluorescence reporter vector (Guo et al., 2015) by Golden Gate assembly (Figure 4.4) and PCR verified using the YGO199 and YGO200 primers (Table of primers) Successfully cloned vectors were transformed into yeast (Figure 4.4). Two clones or each promoter were grown overnight and used for sub-culturing in fresh media to an OD600 of 0.2. The OD600, YFP and mCherry fluorescence were measured on a microtiter plate for 20 hours, with measurements every 10 minutes. The dual fluorescence vector contains an mCherry cassette, driven by pTEF2, as a common element across the entire library (Figure 4.4). Therefore the measured YFP fluorescence is normalised to mCherry to account for extrinsic noise in the cell population across all samples in the library (Sharon et al., 2012). The growth curves and normalised promoter activities were calculated as described in the Materials and Methods section. No visible differences were observed in the growth curves between all strains tested (Appendix figure D.1).

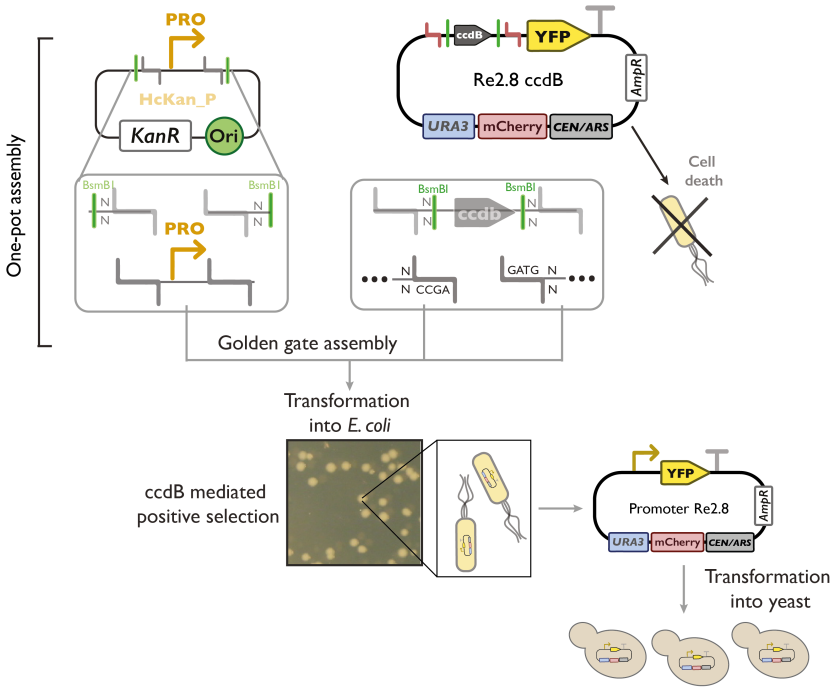


Figure 4.4

An diagrammatic representation of the method used by Guo *et al.* (2015) for cloning promoters into the Re2.8 dual fluorescence reporter vector. The cloning of yeast promoters into the Re2.8 dual fluorescence vector by Golden Gate assembly and subsequently transformed into *E. coli* cells. Positive screening mediated by *ccd*B was used to select for successfully assembled vectors. Plasmids verified for the successful cloning of the promoter into the Re2.8 vector are subsequently transformed into yeast cells.

Normalised activities of each promoter were plotted over time, showing distinct activity profiles. Most promoters show highest activity within the first few hours (between 300-400 mins) of growth (pRPS2, pZEO1, pCYC1, pRPS3, pRPS13) coinciding with the exponential phase of growth. This is followed by a gradual decline starting at 600 mins coinciding with the saturation of the cells in the culture. pTDH3, pACT1 and pPGK1 show a gradual increase in activity within exponential phase, reaching maximum fluorescence production at late exponential phase (∼600 mins) followed by a gradual decline (Figure 4.5A). In contrast, pHSP26 shows a downregulation of fluorescence production throughout exponential phase with a rapid increase in expression towards late exponential phase (∼600 mins) (Figure 4.5A). Maximal activity

was observed during stationary phase (between 800-1000 mins), confirming previous findings by Apel *et al.* about the pHSP26 promoter (Apel et al., 2017).

Similar to previous findings, pTDH3 is the promoter with the highest activity across all the constitutive promoters screened (Apel et al., 2017; Lee et al., 2013; Ottoz and Rudolf, 2018). Comparing the activity of the promoters at exponential and stationary phase, most promoters showed higher activity at exponential phase with the exception of pHSP26, with significantly higher expression in stationary phase (Figure 4.5B). Confirming many previous studies, the stationary phase activity of pHSP26 has been attributed to transcriptional repression mediated by repressing elements on the promoter region (Nussbaum et al., 2014; Susek and Lindquist, 1990; Xiong et al., 2018). Mutations on the pHSP26 promoter by Susek *et al.* has resulted in the constitutive expression of the *HSP26* gene validating the role of these upstream motifs in regulating promoter activity (Susek and Lindquist, 1990). One promoter however showed a different effect to the findings of previous publications. According to the results of Apel *et al.*, pPGK1 appeared to maintain the same level of expression across multiple time points (4 hr, 8 hr, 24 hr and 48 hr). However, the results in this study show that pPGK1 expression gradually increases until late exponential phase, with a maximal fluorescence 2-fold higher than the initial fluorescence measured at the beginning of the growth. Following exponential phase, this activity declines.

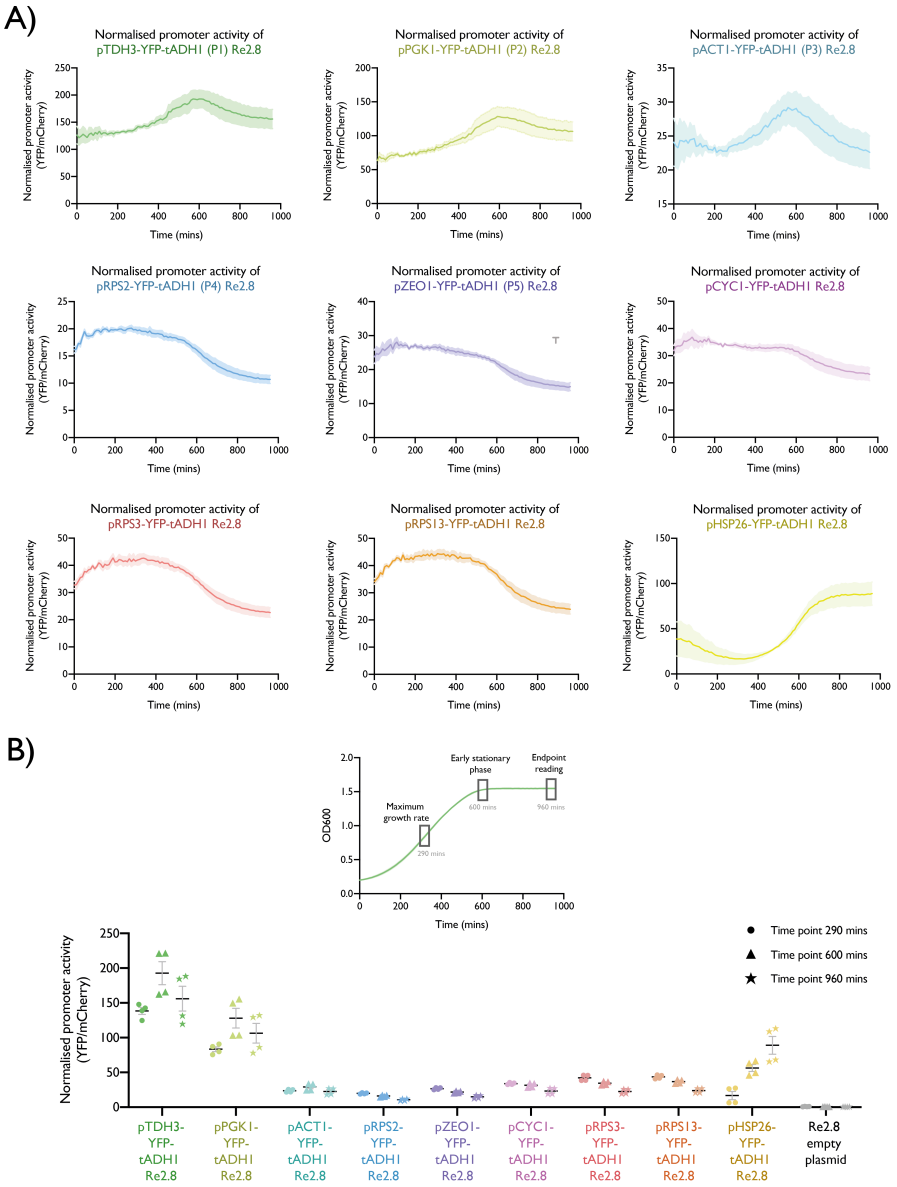


Figure 4.5

A) The characterisation of 9 promoters cloned into the Re2.8 reporter vector. Two clones (biological replicates) for each promoter were analysed for their growth and fluorescence over 24 hours. This assay was repeated three times (n=3). The profiles of each promoter was plotted against. Normalised promoter activities are calculated by normalising the measured YFP fluorescence by the mCherry fluorescence. B) The comparison of promoter activities across all 9 promoters. The normalised promoter activities were taken at 3 time points; at maximum growth rate (290 mins), at early stationary phase (600 mins) and the endpoint reading (960 mins).

One explanation could be due to the limited number of time points measured by Apel et al, missing out time points that may show variations in the expression profile. Comparing the normalised fluorescence measurements recorded in this study at 4, 8 and 20 hours (endpoint), no significant difference in expression was observed, suggesting that the window observed by Apel and colleagues excluded the rise in expression at the ∼10-hour time point (∼600 mins) (Figure 4.5B). This demonstrates the benefit of utilising methods that measure fluorescence production at shorter intervals over 24 hours of growth and how this can reveal information about promoters that would otherwise be missed by a single or four time point screening.

In this section, I have discussed the benefit of utilising the plate reader for obtaining a continuous data set of promoter activity over 24 hours. In the next section, I discuss the adaptation of the protocol for the high-throughput screening of large sets of promoters to identify their promoter activity profile over 24 hours.

4.2.2 High-throughput characterisation of promoter expression profiles over time

The pipeline was adapted for high-throughput growth and screening of a large number of promoters. To this end, 78 promoters were obtained from the YeastFab collection, previously characterised at exponential phase using a flow cytometer. The selection of promoters for further characterisation was based on the following criteria: 1) availability in the collection held at the University Edinburgh, 2) lack of internal BsaI and BsmB1 restriction sites, and 3) diversitiy of promoter activity as measured by Guo *et al.* 2015. Using the characterisation datasets from Guo *et al.* and an unpublished dataset, promoters were ranked based on their relative fluorescence activity (Guo et al., 2015). Promoters containing internal BsaI and BsmB1 sites were excluded to prevent complications in downstream pathway assemblies and only the promoters present in the collection held at the University of Edinburgh were included.

From this list of promoters, the 26 highest expressing promoters were chosen, forming the 'strong promoters' group (S1-S26) (Appendix table D.1). The next 26 strongest promoters (M1-M26) were classed as the 'moderate strength' promoters (Appendix table D.2). The characterisation of promoters by Guo *et al.* showed an asymmetrical distribution of promoters skewed towards weak relative promoter activity. As a result, more than 50% of characterised promoters fell within the promoter activity

range of 0.06-3 arbitrary units. Therefore, for this study, 26 'weak promoters (W1-W26) were chosen in the middle of this range (with promoter activities ranging from 1.2-1.5) (Appendix table D.3).

All promoters were sequence verified prior to subsequent cloning into the Re2.8 reporter vector. In the YeastFab promoter construction pipeline, due to the large number of constructs being made on a high-throughput scale, a tolerance of 4 mutations per promoter were permitted in the final promoter sequence following its cloning into the HcKan_P vector. Therefore, sequence verification of these 78 promoters showed ten promoters with point mutations (Table 4.2). However, all of these were previously identified by Guo *et al.* during the cloning process in the promoter domestication into the HcKan_P vector, possibly introduced during the PCR step.

Four promoters (S16, W8, W25 and W26) out of the 78 screened showed alignment to other promoters, suggesting a possible cross-contamination during the reconstitution of the promoters or in the assembly process (Table 4.2 – Labelled in red). Alignment of the sequencing results to the entire yeast genome showed that the S16_pPXR1 HcKan_P assembly, showed alignment to another promoter, pSCW4 (YGR279C_P). Similarly, the W8_YHL005C_P HcKan_P assembly showed sequencing alignment to the pMRP4 (YHL004W_P) promoter. The sequencing of the W25_YFR032C-B_P HcKan_P showed alignment to the pECO1 (YFR027W_P) promoter. As for W26_pDAP2 HcKan_P, its sequencing result showed alignment to pECM12 (YHR021W-A_P).

Table 4.2 Mutations observed in the sequencing verification of the 78 promoters in the HcKan_P vector. Labelled in red are the promoters with sequencing trace files that have aligned to a different promoter.

Promoter name	Mutations in promoter-HcKan_P plasmids
S2_pENO2	Missing two A's
S9_pPNC1	Insertion of an A and a T within a succession of AT's
S12_pPAL2	Insertion of an A in a succession of A's
S13_pSOD2	Substitution of G to A
S16_pPXR1	Aligns to pSCW4 (YGR279C) promoter
S23_pFUR1	Insertion of an A in a succession of A's
S25_pSOL4	Insertion of an T in a succession of T's
M8_pGRE3	Deletion of one T
M20_pNIC96	Deletion of AATA
W6_pHSN1	Insertion of an A in a succession of A's
W8_YHL005C_P	Aligns to another promoter (YHL004W)
W22_YHR138C_P	Loss of a T in a succession of T's
W25_YFR032C-B_P	Aligns to YFR027W promoter
W26_pDAP2	Aligns to YHR021W-A promoter

The remaining promoters were cloned into the Re2.8 vector by Golden Gate assembly as described in Figure 4.4 and verified for their successful cloning into the Re2.8 reporter vector by PCR using YGO199 and YGO200 (Table of primers) (Appendix Figure D.2A). All assemblies showed the correct size of amplicon for all 78 promoters and were subsequently transformed into wild type yeast (BY4741). However, as four promoters in the collection previously showed sequencing alignment to other promoters, they were excluded from subsequent analysis. Biological replicates for each promoter reporter strain were inoculated on fresh media and grown overnight on a 96-deep well plate using a protocol for yeast culturing previously optimised by Dr. Eva Garcia-Ruiz. The strains were diluted to an OD of 0.2 prior to the plate reader-based measurement of growth and fluorescence over 24 hours. Measurements of growth and fluorescence for all 74 promoters were repeated three times (n=3). Growth curves across all three experimental repeats showed different lengths in lag time. Therefore, all growth and fluorescence curves for all 74 promoters across the three replicates were normalised by the peak growth rate (Materials and Methods section). Post normalisation no visible differences were observed in the growth curves across all 74 promoters (Appendix Figure D.3-D.5).

The promoter activities from the Guo *et al.* publication was compared to the results from this plate reader assay to determine if similar rankings between the 74 promoters were obtained. To this end, the promoter activities were normalised in a similar fashion to that carried out by Guo *et al.* The normalised YFP fluorescence (YFP/mCherry) of all 74 promoters at OD600 of 1 were extracted and divided by the normalised YFP fluorescence (at OD600 of 1) of pCYC1. Guo *et al.* carried out this normalisation to be able to rank the promoters based on their promoter activity relative to pCYC1 (Figure 4.6A). Comparison between the two datasets showed a similar trend to that of Guo *et al.*, with most strong promoters showing the highest expression among all 74 promoters and the weak promoters showing the lowest relative promoter activities (Figure 4.6B). Differences in the y-axis between the two graphs can be attributed to the different method of characterising promoter activity (flow cytometry vs. plate reader). Several promoters did not conform to the trend. Two strong promoters (S5_pOCH1 and S10_pRPL22B), and four moderate strength promoters (M1_pAAD6, M2_pDIA4, M6_pBCD1 and M14_pPAU13) showed relative promoter activities significantly lower than any weak promoter characterised in this collection. Multiple factors may contribute to this difference such as the culturing conditions (growth in 1 ml media in a 96 deep well plate for 8 hours on a plate shaker vs. growth in 200 μ L of media on a microtiter plate for 24 hours in a plate reader) and analysis (flow cytometry vs. plate reader). Analysis of large libraries of promoter are currently based on the characterisation at one time point, which varies between datasets. These promoter activities are then used to rank the promoters. However, they only represent the promoter's activity at one time point. The comparison carried out in (Figure 4.6A and B) highlights that different methods of analysing the same strains can yield different promoter activities, ultimately affecting its 'activity ranking' within a library. Given the variability observed when sampling at single time points, I sought to dissect the dynamics of expression over time.

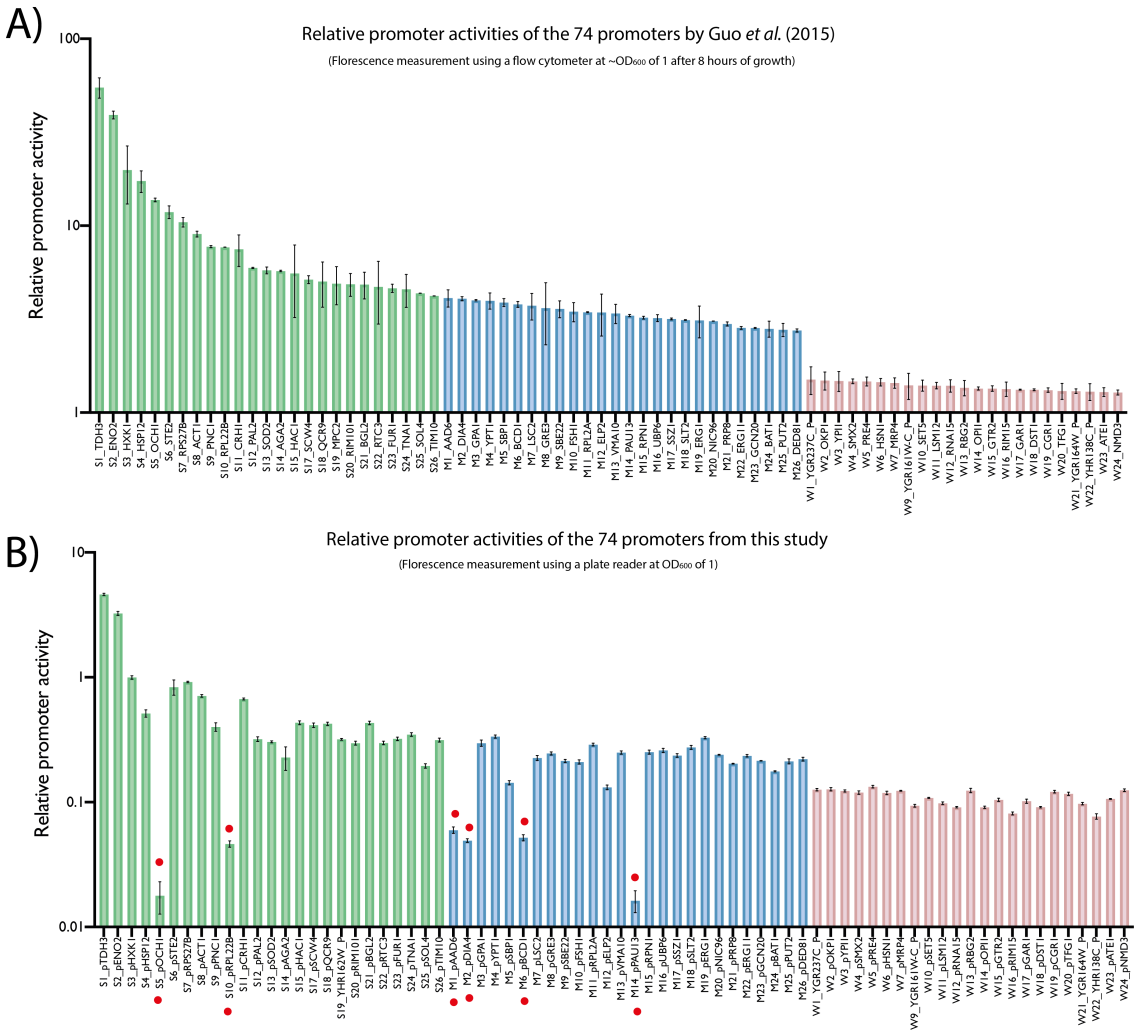


Figure 4.6

Comparison of relative promoter activities of 74 promoters between the Guo *et al.* dataset and results obtained in this study. Promoters in green classified as 'strong' promoters, promoters in blue classified as 'moderate strength' promoters, and in pink classified as 'weak' promoters. A) Relative promoter activities of 74 promoters (relative to the promoter activity of pCYC1) in the YeastFab collection, characterised by Guo *et al.* using a flow cytometer (Guo et al., 2015). B) Relative promoter activities of 74 promoters (relative to the promoter activity of pCYC1) in the YeastFab collection, characterised in this study using the plate reader. (Marked with red dots are promoters with significantly lower promoter expression relative to the data obtained by (Guo et al., 2015))

Figure 4.7-4.9 show the activity profiles of the library of promoters over time. The YFP fluorescence was normalised to the mCherry fluorescence, driven by pTEF2 promoter to account for extrinsic noise in the cell population across all samples in

the library (Sharon *et al.*, 2012). Most promoters display a decrease in normalised YFP fluorescence for the first ∼200 mins, before showing increasing fluorescence until ∼450 mins as the population enters stationary phase (Figure 4.7-4.9). After peaking at ∼450 mins the fluorescence shows a gradual decline in activity for the remainder of the measurement window (Figure 4.7-4.9). These results show that majority of promoters show maximal activity at late exponential phase.

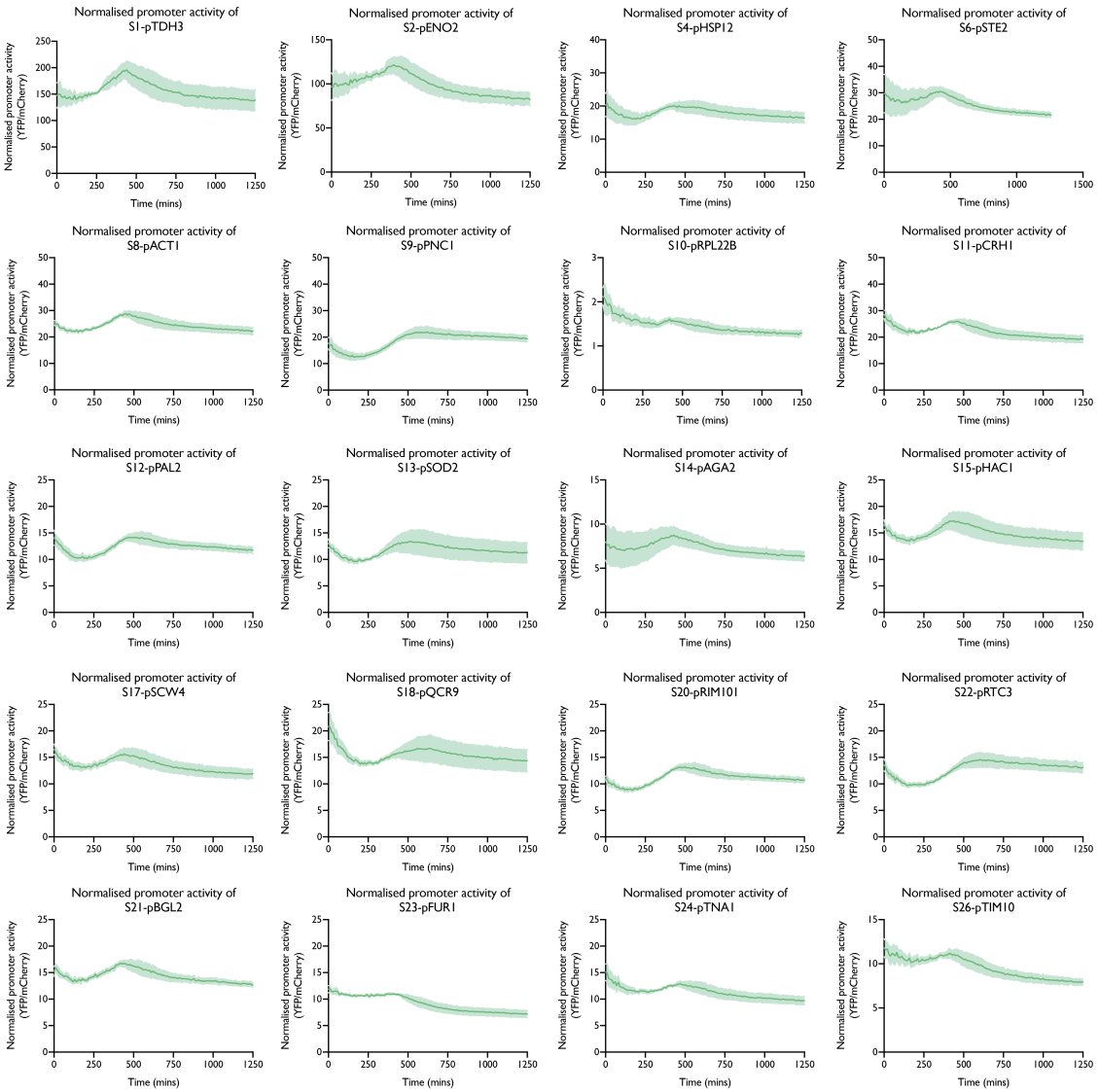


Figure 4.7

Characterisation of 'strong' promoters over time (n=3). The normalised promoter activity was calculated by dividing the blank corrected YFP fluorescence by the blank corrected mCherry fluorescence. In order to account for differences in the lag time, the fluorescence measurements for each promoter was normalised to the maximum growth rate across all replicates (Appendix figure D.3) – for OD600 curves)

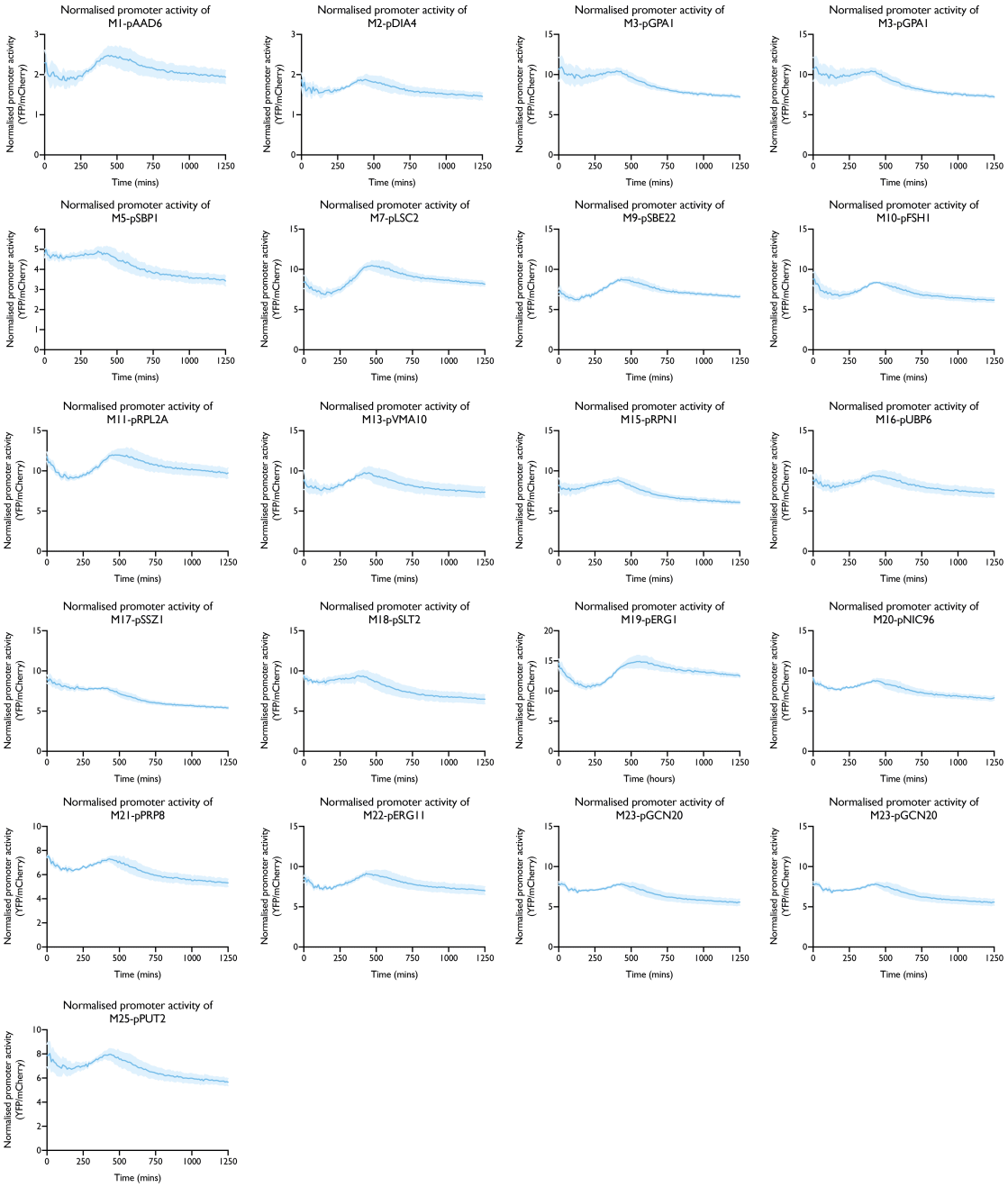


Figure 4.8

Characterisation of 'moderate strength' promoters (n=3). The normalised promoter activity was calculated by dividing the blank corrected YFP fluorescence by the blank corrected mCherry fluorescence. In order to account for differences in the lag time, the fluorescence measurements for each promoter was normalised to the maximum growth rate across all replicates (Appendix figure D.4 – for OD600 curves)

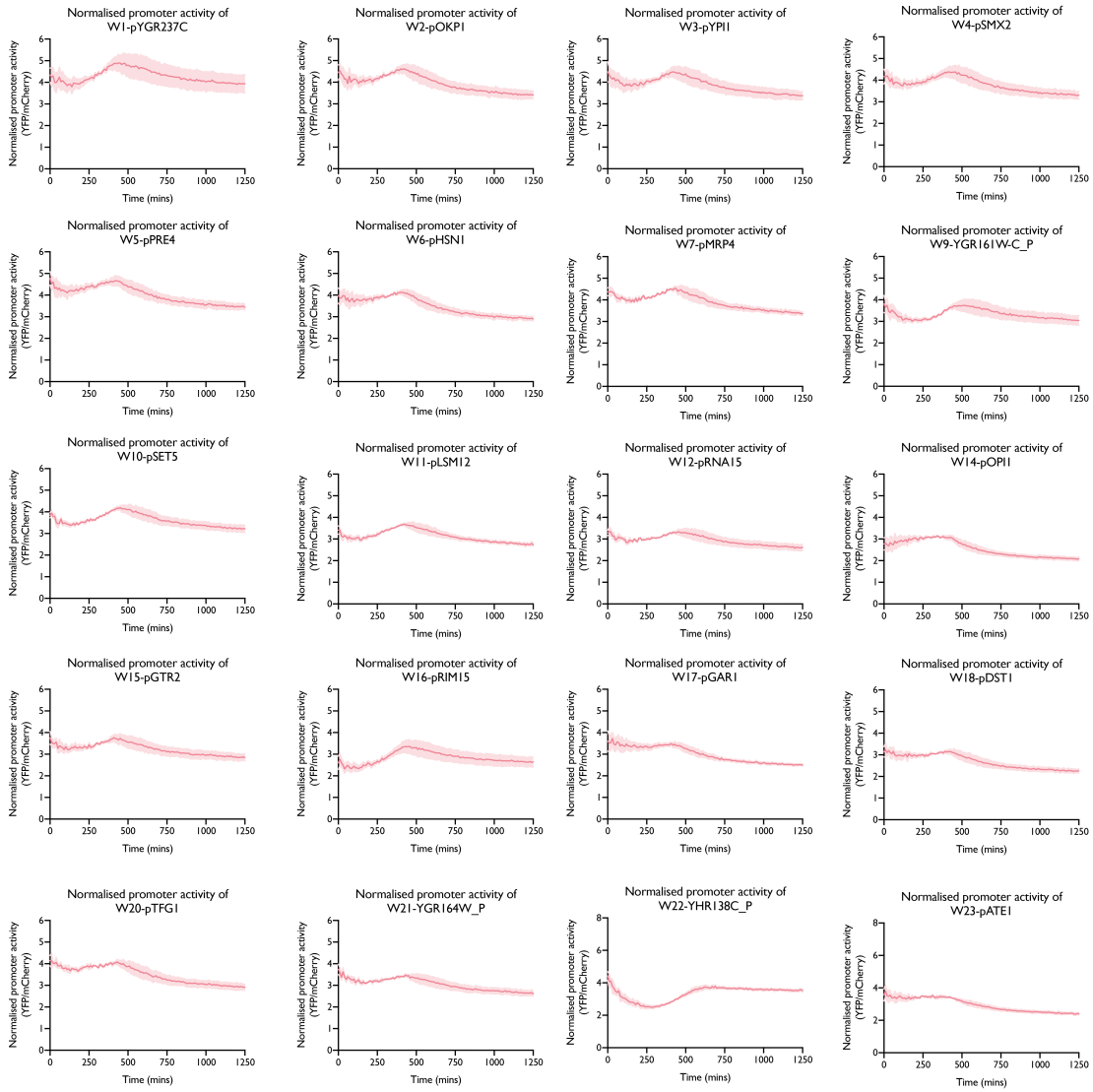


Figure 4.9

Characterisation of 'weak strength' promoters (n=3). The normalised promoter activity was calculated by dividing the blank corrected YFP fluorescence by the blank corrected mCherry fluorescence. In order to account for differences in the lag time, the fluorescence measurements for each promoter was normalised to the maximum growth rate across all replicates (Appendix figure D.5 – for OD600 curves)

In contrast, several promoters appear to function as mid-exponential phase promoters, showing high activity in early growth (the first ∼6 hours of growth). pRPS27B (S7), pOCH1 (S5), pBCD1 (M6), pPAU13 (M14), pRBG1 (W13), pCGR1 (W19) and

pNMD3 (W24) follow a similar trend to that observed for ribosomal protein associated promoters screened in the previous section (pRPS3, pRPS2, pRPS13)(Figure 4.10). Annotated functions of the corresponding genes of the promoters above are involved in transcription or translation machinery, appropriately reflecting their maximal activity within early to mid-exponential phase with a gradual decline towards late exponential to stationary phase.

Other exceptions of note include pHXK1 (S3) and pGRE3 (M8) which show maximal activity in stationary phase (Figure 4.11). pHXK1 has been previously shown to be responsive to the presence of glucose, increasing in expression with the depletion of glucose. This function appears to be reflected by the promoter activity profile observed in this study.

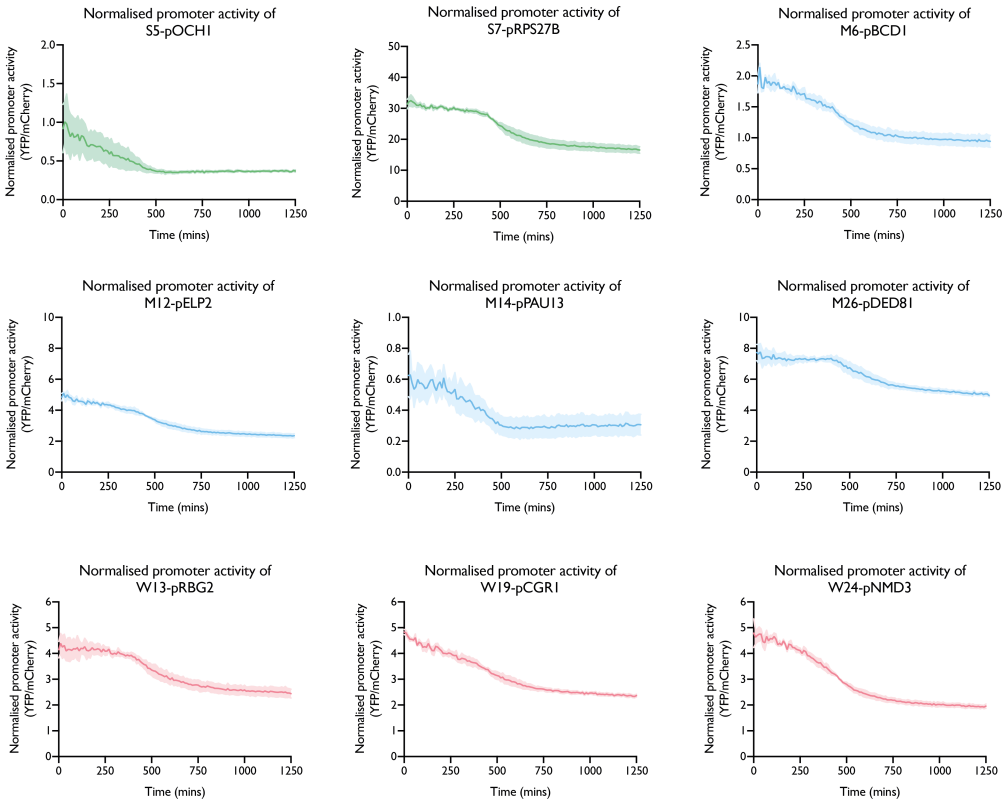


Figure 4.10

Promoters showing decline in normalised YFP expression throughout the culturing period, with peak activities within exponential phase. The normalised promoter activity was calculated by dividing the blank corrected YFP fluorescence by the blank corrected mCherry fluorescence. In order to account for differences in the lag time, the fluorescence measurements for each promoter was normalised to the maximum growth rate across all replicates (Appendix figure D.3-D.5 – for OD600 curves)

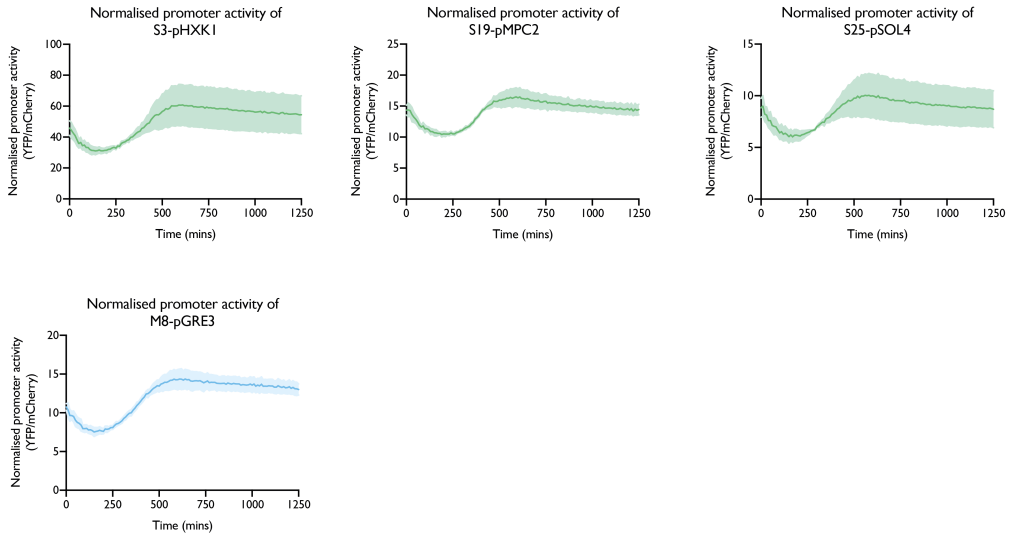


Figure 4.11

Promoters with activities showing a drop in YFP fluorescence expression in exponential phase and an increase upon saturation of the culture. The normalised promoter activity was calculated by dividing the blank corrected YFP fluorescence by the blank corrected mCherry fluorescence. In order to account for differences in the lag time, the fluorescence measurements for each promoter was normalised to the maximum growth rate across all replicates (Appendix figure 3-5 – for OD600 curves)

The graphs highlight that for most promoters screened, the variation between the expression over 20 hours of culturing is low (Figure 4.7-4.9). As such many of the promoters characterised here can be expected to function as constitutive promoters. However, we identify some outliers, which show much higher variability, either showing significantly increased or decreased expression over the course of culturing. As such for sensitive applications, where expression levels over the course of culturing may significantly alter titres or toxicity, it would be sensible to utilise this or a comparable screening pipeline to ensure a desired expression profile.

4.2.3 Characterisation of the Z3EV transactivation system

A higher level of gene expression control, in particular of key enzymes in a metabolic pathway can be mediated using regulated (inducible or repressible) systems. A widely used inducible promoter for overexpression of genes is the GAL system.

However as previously described, utilisation of this system requires growth in media containing alternative carbon sources such as raffinose. Sadowski modified to system by fusing the GAL4 DNA binding domain to to VP16, making it a potent transactivator while allowing growth in glucose containing media (Sadowski et al., 1988). Louvain *et al.* further modified the system by introducing a β-estradiol binding domain (EBD) allowing this system to conditionally activate downstream gene expression in the presence of estradiol (Louvion et al., 1993). This transactivator was therefore called GEV for Gal-EBD-VP16. However, due to multiple potential binding sites of the Gal DBD of the GEV system, McIsaac *et al.* created an orthogonal system, the Z_3EV system, with fewer predicted binding sites in the yeast genome. The Z_3EV system, consists of a three fingered Cys2His2 zinc-finger DBD, EBD and a VP16 activator (Figure 4.12A). The zinc finger DBD target a Cys2His2 zinc-finger (Zif268) binding motif (GCGTGGGCGT), which was were used to replace the Gal4 binding motifs in the UAS of the pGAL promoter. Without induction with β-estradiol, the estradiol receptor (ER) on the Z_3EV interacts with the Hsp90 chaperone complex which sequesters the protein complex to the cytoplasm. Induction with β-estradiol results in the displacement of Hsp90 by β-estradiol and conformation change of the ER on the Z_3EV protein, exposing a nuclear localisation signal (Figure 4.12A). This targets the Z_3EV protein to the nucleus where the zinc finger-based DNA binding domain binds to its corresponding motif. The VP16 domain in the Z_3EV complex can subsequently recruit multiple factors that play a role in transcription activation; TATA binding protein (TBP), transcription factors *TFIIIA*, *TFIIB*, *TFIIF* and *TFIIH*, as well as interacting directly with the mediator complex (Hirai et al., 2010; Sadowski et al., 1988; Uhlmann et al., 2007). The following body of work consists of the domestication of pZ3EV into the YeastFab format and the characterisation of its expression profile over time, upon induction with β-estradiol.

The Z_3EV system consists of two components, the Z_3EV transcriptional activator and the pZ3EV promoter consisting of a modified pGAL1 promoter with the Gal4 DNA binding domains replaced with 6x Cys2His2 zinc-finger (Zif268) binding motifs (GCGTGGGCGT), with the last base showing low sequence conservation (Figure 4.12B and C). The cassette encoding the Z_3EV activator was integrated into the *TRP1* locus by collaborators (Susan Moore) at the University of Montreal. This strain construction was carried out by replacing the endogenous *TRP1* promoter, ORF and terminator with the pACT1 and tENO2 driving the expression of Z_3EV (Appendix figure D.6). The resulting strain was subsequently called SASy31 for the MATα strain and SASy35 for the MATa strain with both possessing the Z_3EV transactivator constitutively expressed.

The constitutive expression of Z_3EV does not result in the upregulation of the target gene due to the localisation of the Z_3EV in the cytoplasm, preventing targeting of the genome.

The pZ3EV promoter was domesticated to the YeastFab standard allowing easy cloning of the promoter upstream of the gene of interest. The promoter was amplified using McIsaac-HcK_FP and McIsaac-HcK_RP (Table of primers) from a plasmid previously constructed by our collaborator (pSASS1). BsaI and the appropriate prefix and suffix 4 bp overhangs were incorporated through the designed primers, enabling subsequent Golden Gate assembly of the promoter into the HcKan_P vector. The promoter was successfully constructed, and sequence verified (Figure 4.12B).

Further analysis of the sequencing results revealed a mutation within the Zif268 binding motifs (Figure 4.12B). The mutation was in the last base in one of the repeats of the 6x Zif268 binding motifs with low sequence conservation (McIsaac et al., 2013). Comparison of this promoter sequence with the original promoter design did not show consistency in the last base used for the Zif268 binding motifs (combination of T's and G's) showing that the last base was not significant in the functioning of the Z_3EV activator (Figure 4.12C) (McIsaac et al., 2013).

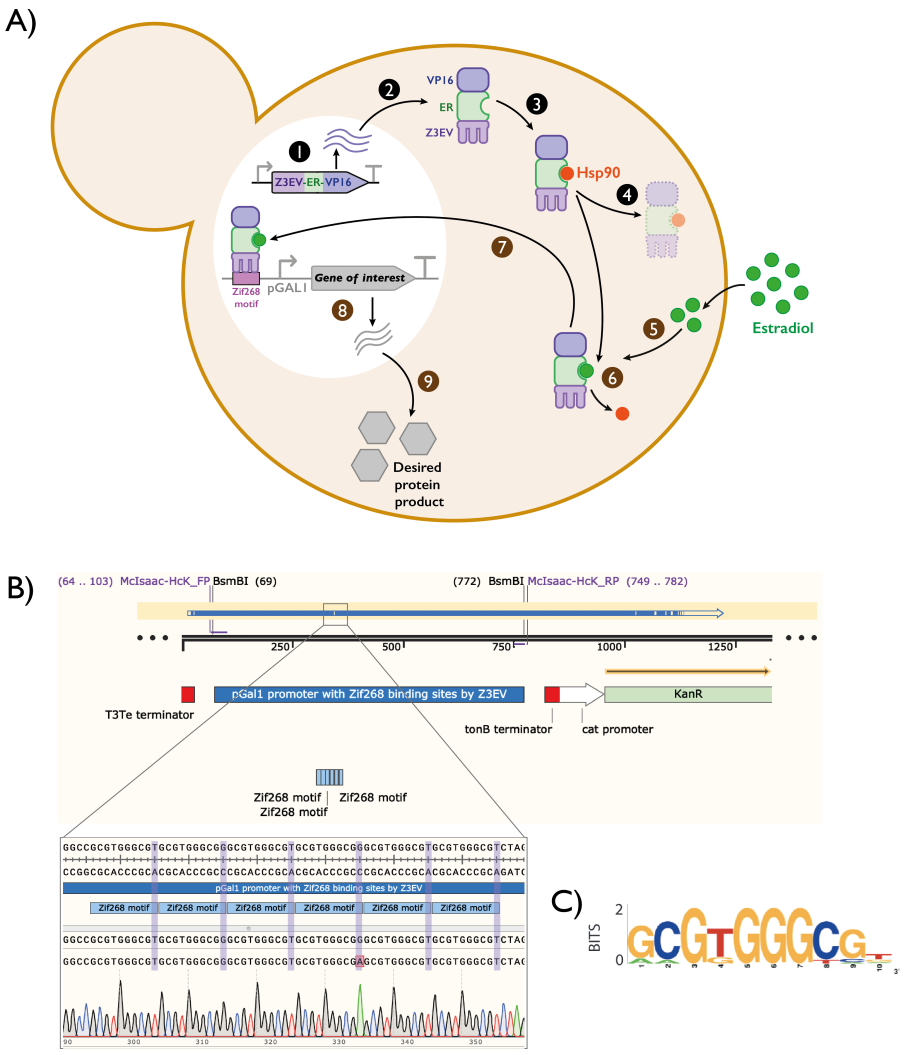


Figure 4.12

Overview of the Z_3EV inducible system (McIsaac et al., 2013). A) Without induction with β-estradiol, the estradiol receptor (ER) on the Z_3EV interacts with the Hsp90 chaperone complex which sequesters the protein complex to the cytoplasm. Induction with β-estradiol results in the displacement of Hsp90 by β-estradiol and conformation change of the ER on the Z_3EV protein, exposing a nuclear localisation signal. This targets the Z_3EV protein to the nucleus where the zinc finger-based DNA binding domain binds to its corresponding motif. The VP16 domain in the Z_3EV complex can subsequently recruit multiple factors that play a role in transcription activation (McIsaac et al., 2013). B) Sequencing verification of the domesticated pZ3EV promoter showed a mutation on the 4th Zif268 motif, a substitution of a G to an A. However, comparison of the last base on the Zif268 motif across all 6 repeats shows no consistency, with several using a T while others use a G (labelled with a purple box). C) Binding motif sequence conservation from the McIsaac *et al.* (2013) publication, showing low sequence conservation of the 10th base of the Z_3EV binding motif (McIsaac et al., 2013).

This promoter was subsequently cloned into the YeastFab Re2.8 reporter vector, upstream of the YFP cassette. Following PCR verification, the resulting plasmid was transformed into the SASy31 strain for subsequent characterisation. Using a plate reader, the fluorescence and OD were measured in the pZ3EV Re2.8 SASy31 strain induced with different concentration of β-estradiol. The growth of the pZ3EV Re2.8 SASy31 strain (JAy509) was compared to that of BY4741 containing the empty Re2.8 vector (JAy008).

Analysis of the OD600 curves between the two background strains (BY4741 vs. SASy31) without the inducer showed a decrease of the growth at late exponential phase for the SASy31 background (JAy509) (Figure 4.13A). This decrease is however not observed in strains with a BY4741 background (JAy008) regardless of the presence of β-estradiol (Figure 4.13B), suggesting that the growth difference is not due to β-estradiol toxicity. Comparison of the growth curves of JAy509 strain induced with concentrations of β-estradiol ranging from 0.25 μM to 2 μM (Figure 4.13B) showed the same growth pattern as JAy509 with no inducer (Figure 4.13A – red line). No difference in growth pattern was observed with increasing concentrations of β-estradiol, confirming previous findings by Mcsaac *et al.* (McIsaac et al., 2013).

Of importance, there is a difference in growth patterns between the background strains (SASy31 vs. BY474) irrespective of the induction or concentration of β-estradiol (Figure 4.13B) showing that this is not due to inducer-dependent toxicity. This decrease in growth can be attributed to the differences in background strains such as the constitutive expression of the Z_3EV transactivator activator or the location of Z_3EV knock-in resulting in a trp- strain. In order to account for this difference in growth, the measured fluorescence was normalised to their respective OD600 values for each sample.

The normalised fluorescence curves of JAy008 induced with β-estradiol show a bimodal peak. One occurring in early exponential phase another in early stationary phase. This pattern has not previously been observed. It is hypothesised that upon introduction of the inducer (0 mins), the transcription rate is very high and whilst the growth rate is low, the protein per cell (measured by YFP/OD) increases. However, as the growth rate increases the dilutive effect on the protein supersedes the role of the high transcription rate leading to a relative decline in YFP. As the population enters stationary phase the growth rate diminishes, and we initially see an increase in YFP relative to OD600 as the dilutive effect on YFP per cell recedes (Figure 4.13). Finally,

at later stationary phase we see a steady decline in protein abundance, similar to that observed in promoters characterised in Figure 4.7-4.9).

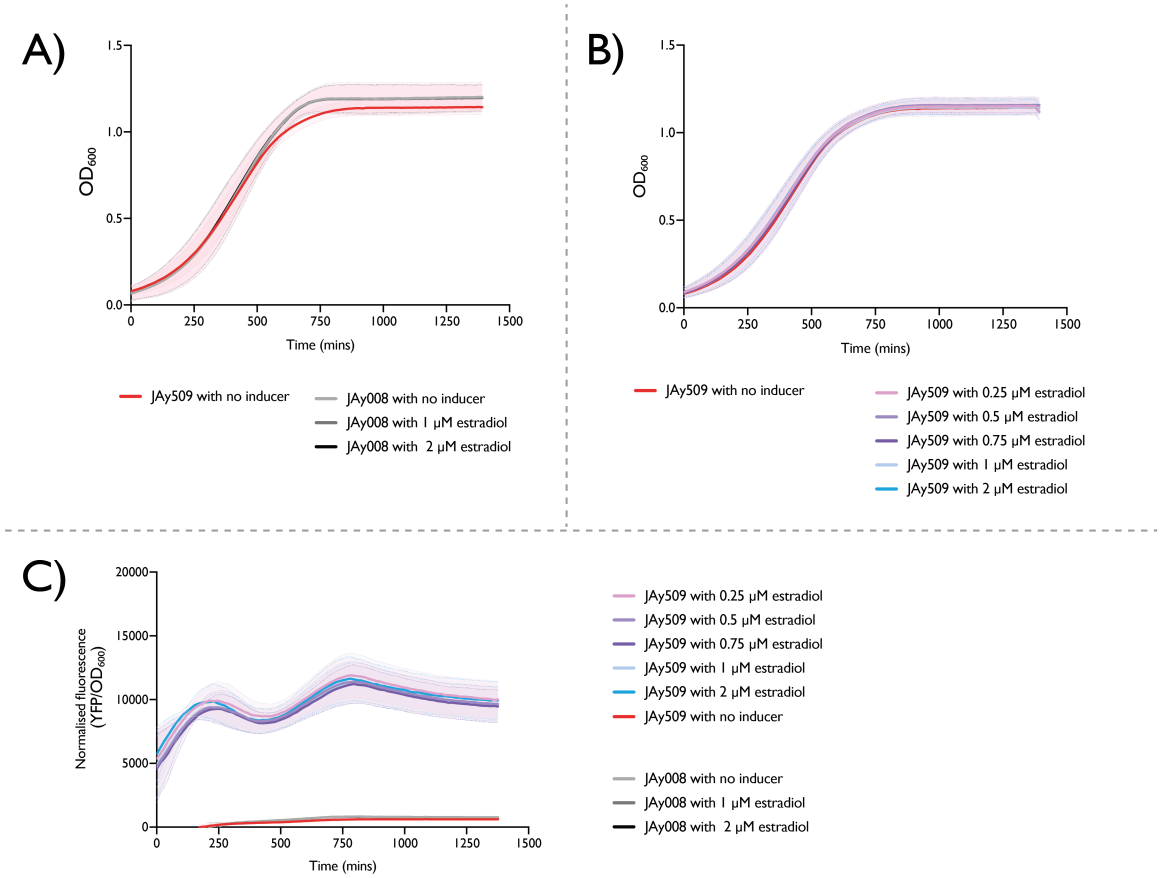


Figure 4.13

A) Comparison of growth curves between JAy509 (a yeast strain with the pZ3EV cloned into the Re2.8 vector) and JAy008 (a yeast strain transformed with an empty Re2.8 vector) without the addition of the inducer. JAy509 and JAy008 have different background strains, SASy31 and BY4741 respectively. The growth curve JAy008 with the addition of the inducer was also tested to determine any toxicity effects of β-estradiol. B) Growth curves of JAy509 (a SASy31 yeast strain transformed with the pZ3EV cloned into the Re2.8 vector) induced with different concentrations of β-estradiol (0.25μM-2μM). C) Comparison of the normalised YFP fluorescence of JAy509 and JAy008 induced with different concentrations of β-estradiol.

4.2.4 Characterisation of the neomycin riboswitch system

Another example of a regulatable systems are riboswitches. Protein production can be controlled post-transcriptionally, by engineering RNA to function as molecular switches or riboswitches that can prevent translation of an mRNA in the presence of a ligand. Aminoglycosides are a group of broad-spectrum antibiotics that act through interruption of protein synthesis by binding to the A-site of bacterial 16S rRNA. Structurally, aminoglycosides have a chemical characteristic of high flexibility due to the presence of sugar rings and amine groups. This flexibility aids interaction with the RNA binding pocket within internal loops of the RNA helix. In addition, the positive charges on the molecule are attracted to the negatively charged RNA backbone aiding the interaction between the small molecule and the aptamer. All these properties make aminoglycosides a tractable group of compounds to design aptamers against.

Weigand and colleagues successfully identified and optimised an aptamer that regulated gene expression in the presence of neomycin (Figure 4.14A) (Weigand et al., 2008). Efficiency of the riboswitch is determined by the switching factor, which is defined as the fold-change of GFP fluorescence for the de-repressed compared to the repressed conditions. They observed that gene expression of the reporter was regulated in a dose dependent fashion, with at least 800 nM of neomycin required for efficient switching. Specificity studies show that the neomycin aptamer is not responsive to a closely related aminoglycoside, paromomycin which differs only by a substitution of the amino group with a hydroxyl group at position 6 in ring 1. However, mild repression was observed when cells were treated with an analogue of neomycin, ribostamycin which lacks ring IV (Weigand et al., 2008).

In the following work, I sought to investigate the activity of neomycin within the context utilised for the pathway expression and compound production in yeast. These include its expression in BY4741 background and from low copy plasmids. To this end, the pWHE601 vector containing the M4 neomycin riboswitch was provided by the Suess lab (Figure 4.14A). This plasmid contained the neomycin M4 aptamer cloned within the 5'UTR immediately upstream of the start codon (Figure 4.14A). The pADH1 promoter controlled the expression of the GFP reporter expressed from 2μ multi-copy vector. This plasmid was transformed into wild type BY4741 yeast and selected in Sc -Ura dropout media to select for the pWHE601 plasmid maintenance. The resulting strain, JAy109, was further characterised using the plate reader.

Different concentrations of neomycin ranging from 5 μM to 1 mM were introduced into the yeast medium at the beginning of the plate reader growth assay. The fluorescence and absorbance at OD600 were measured over 24 hours. Decrease in growth was observed with increasing concentrations of neomycin added, in particular for 1 mM neomycin (Figure 4.14B). This suggests a dosage dependent toxicity of neomycin for *S. cerevisiae*, consistent with previous reports (Weigand et al., 2008).

Fluorescence measurements were normalised to OD600 to account for the variability in the data introduced due to differences in growth. A media correction was also carried out, involving the subtraction of the fluorescence produced from the media in the fluorescence measurements across all samples. Despite this correction, a baseline GFP fluorescence was observed even in the sample with a high concentration of neomycin (1 mM) added (Figure 4.14C). This suggests a possible 'leaky' expression however further characterisation studies need to be carried out to verify this result.

Repression of translation mediated by the riboswitch was observed in all JAy109 samples with neomycin added relative to the sample with no ligand addition (Figure 4.14C). At least a three-fold decrease in normalised fluorescence was observed in JAy109 with the lowest concentration of neomycin added (5 μM) (Figure 4.14D). In alignment with data published by Weigand *et al.*, the fluorescence decreases with increasing concentrations of neomycin (Figure 4.14D) (Weigand et al., 2008). The switching factor was calculated using the endpoint normalised fluorescence for all samples induced with neomycin. With increasing concentration of neomycin, up to 1 mM, a higher switching factor was observed showing a more efficient repression of translation (Figure 4.14D). The switching factor for 1 mM showed almost 2x the switching factor of the lowest concentration tested (5 μM). An upward trend was observed in JAy109 with increasing concentrations of neomycin added, suggesting that the increase in the ligand resulted in an improved switching factor. To test this, a statistical test for linear trend was performed, showing a significant upwards trend for JAy109 with an R^2 value of 0.9565 ($P<0.0001$). However, a growth defect was observed with the use of higher concentrations of neomycin, especially with 1 mM.

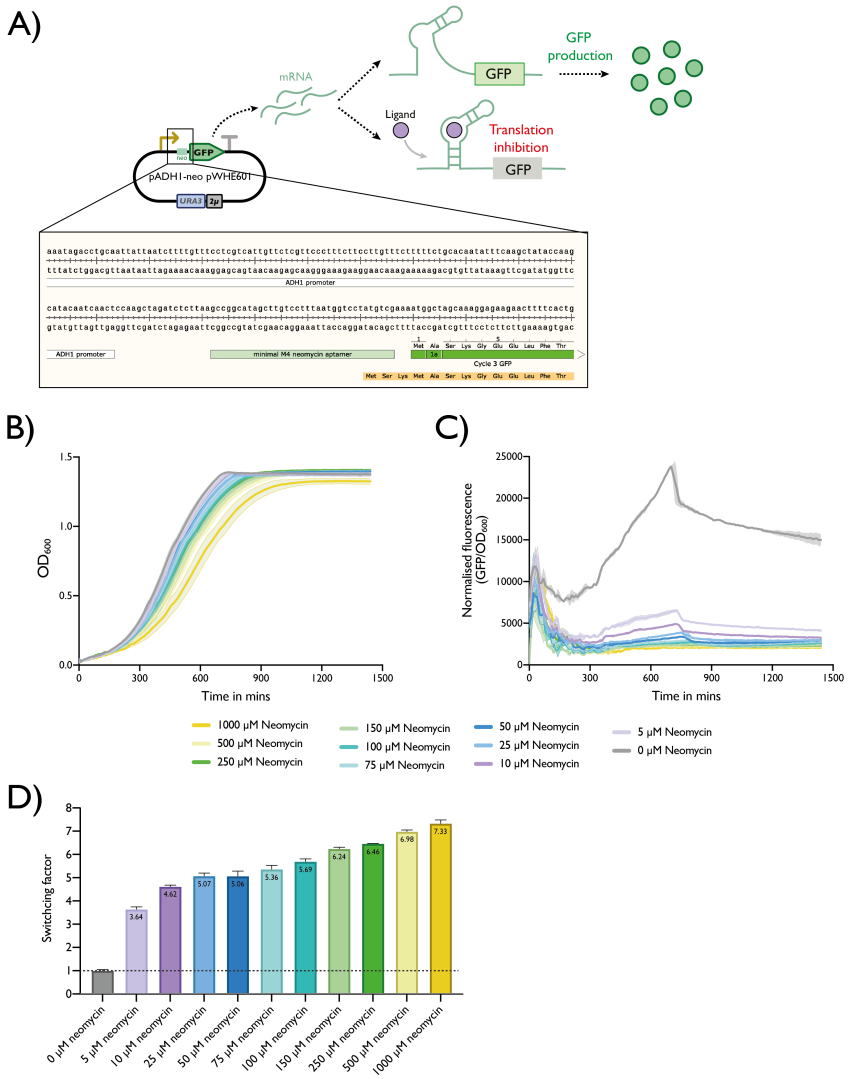


Figure 4.14

A) An overview of the mechanism of riboswitch-based repression of translation. The minimal N4 neomycin aptamer lies within the ADH1 5'UTR, immediately upstream of the start codon. In the unbound state the aptamer possesses a weak secondary structure in the 5'UTR. Binding of the ligand to the aptamer, in this case neomycin, results in the formation of a stable secondary structure in the 5'UTR resulting in translation inhibition. B) Growth curves for JAy109 (BY4741 yeast strain containing the neomycin pWHE601 reporter plasmid) grown with different concentrations of neomycin. C) GFP fluorescence normalised to OD600 of JAy109 grown with different concentrations of neomycin. D) Switching factor of the JAy109 added with concentrations of neomycin. Switching factor is calculated by dividing the endpoint fluorescence of the strain with no neomycin added by the fluorescence of each strain grown with neomycin.

4.2.5 Domestication of the neomycin riboswitch to YeastFab standards

To test the activity of the riboswitch in a low copy plasmid, primers were specifically designed to amplify both the pADH1 promoter and the M4 neomycin riboswitch. Within these primers, BsmBI restriction sites, the correct overhangs and 40 bp homology to the HcKan_P vector were incorporated to the promoter-riboswitch sequence (Figure 4.15A). This PCR product was size verified and cloned into the HcKan_P vector by Gibson assembly. Sanger sequencing verification of a clone from this assembly showed the successful assembly of the promoter into the HcKan_P acceptor vector (Figure 4.15B). Only 1 mutation was found in the pADH1 promoter; a deletion of an A in a succession of A's. However, sequencing of the pADH1 promoter on the original pWHE601 vector also showed the same modification (Figure 4.15C). As this mutation was not within the 5'UTR of pADH1 or the neomycin aptamer and found in the published plasmid by the Suess lab, we decided to proceed with the cloning and characterisation using the current version. The pADH1-neo riboswitch was cloned into the Re2.8 dual fluorescence vector by Golden Gate assembly as previously described in Figure 4.4. Sequencing verification showed successful cloning of the riboswitch into the Re2.8 reporter vector, with the same deletion of an A in a succession of A's as previously found (Figure 4.15D).

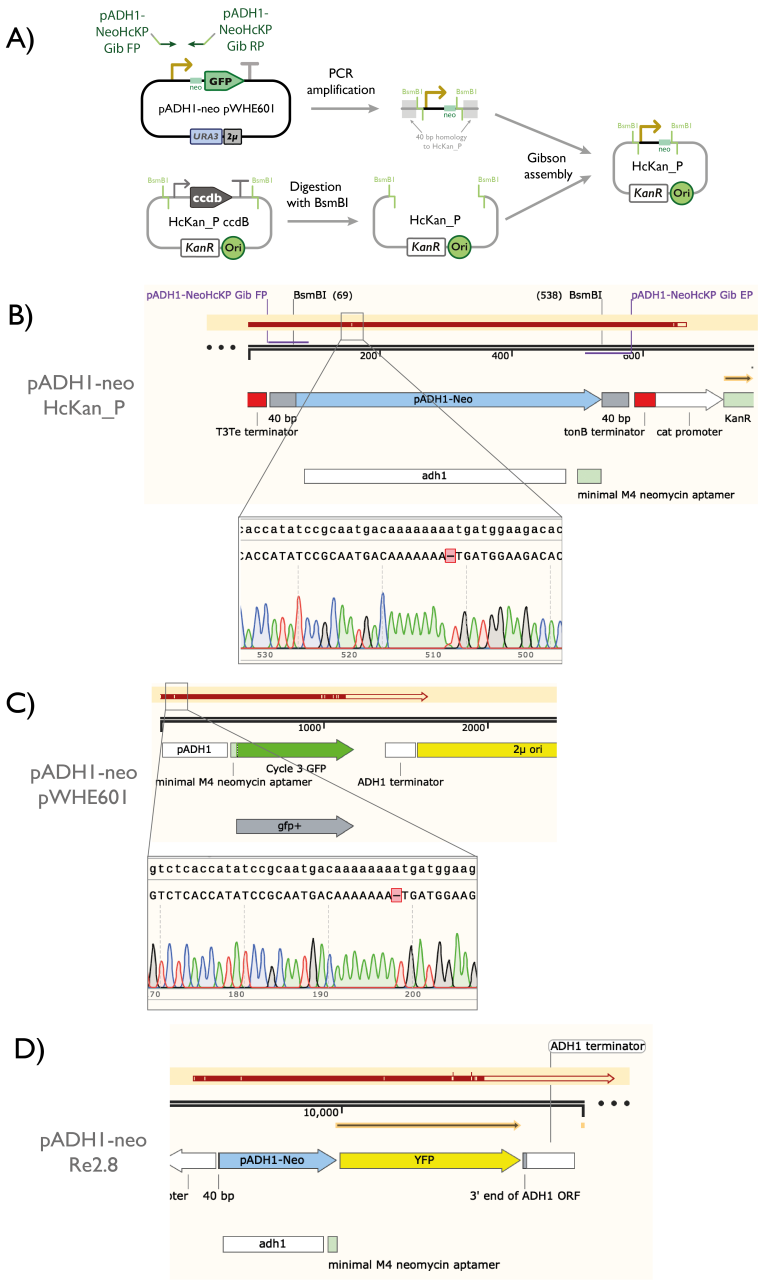


Figure 4.15

A) Cloning process for the domestication of the neomycin riboswitch to the Hckan_P vector. Primers were used to incorporate a 40 bp homology region to the HcKan_P vector flanking the pADH1-neo promoter. The pADH1-neo promoter was assembled into the HcKan_P vector by Gibson assembly. B) Sequencing verification of pADH1-neo HcKan_P, showing a deletion of one A in a succession of A's in the pADH1 promoter. C) Sequencing verification of the pADH1-neo pWHE601 plasmid from the Suess lab, showing the same deletion of one A in the pADH1 promoter. D) Sequencing verification of pADH1-neo in the Re2.8 dual fluorescence reporter vector.

The JAy508 strain (BY4741 containing the pADH1-neo Re2.8) was grown in a plate reader with different concentrations of neomycin (0-150 μM). Slight differences in the growth curves of the strains were observed, supporting the dosage dependent growth inhibition observed for previous results (Figure 4.16A).

The fluorescence measured for JAy508 was normalised to OD600, instead of mCherry, similar to the previous characterisation of JAy109 (BY4741 with the pADH1-neo pWHE601 plasmid). This was to account for the variation in growth of JAy508 treated with different concentrations of neomycin. The background YFP fluorescence was subtracted for all samples however despite this, a baseline fluorescence was observed across all neomycin treated samples, similar to that observed with JAy109.

Similar to previous findings, all JAy508 samples added showed approximately 3-fold lower YFP production compared to the sample with no neomycin added. Increasing neomycin concentrations showed a slightly lower YFP production showing an inverse correlation between protein production and amount of neomycin added to the media (Figure 4.16B). The endpoint normalised fluorescence was used to calculate the switching factor in JAy508 with different concentrations added. When performing a Dunnett's multiple comparison test, all JAy508 samples show a significant difference in the switching factor upon the addition of neomycin across all concentrations tested compared to the no ligand control ($P<0.0001$ for all). Similar to previous findings in (Figure 4.15D), increasing amounts of neomycin resulted in an upwards trend in switching factor (Figure 4.16C). Analysis of JAy508 with different concentrations of neomycin using a statistical test for linear trend showed a significant upwards trend for JAy508 ($R^2=0.8246$) ($P<0.0001$).

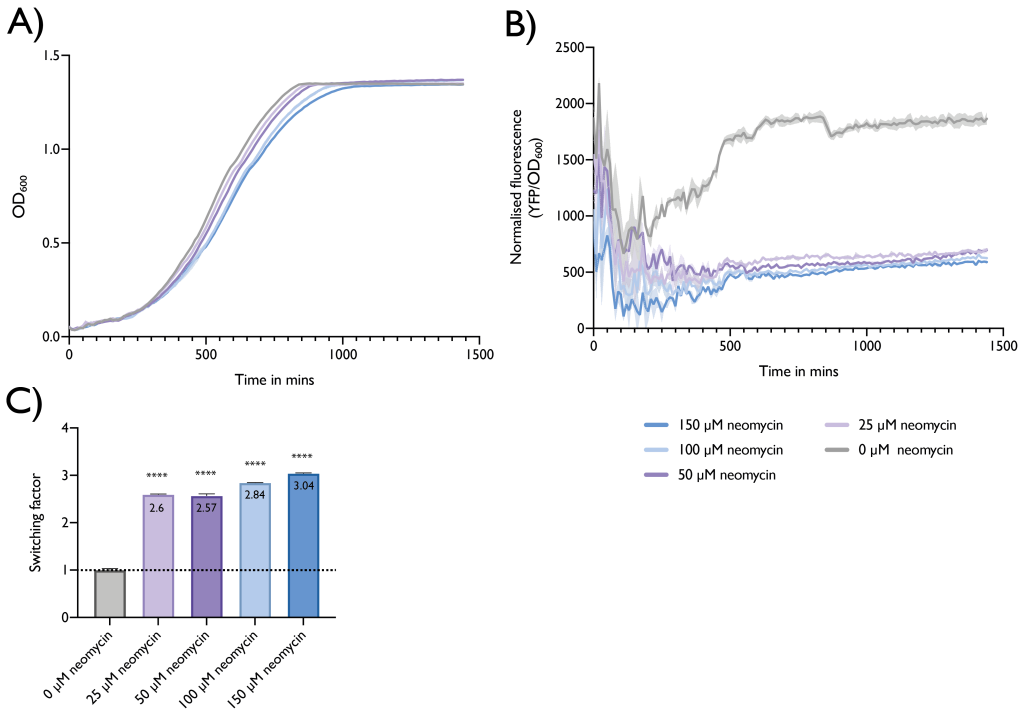


Figure 4.16

Characterisation of the neomycin riboswitch expressed from a low copy CEN/ARS plasmid. A) JAy508 grown in different concentrations of neomycin (0-150 μM). B) YFP fluorescence of JAy508 grown in different concentrations of neomycin, normalised to OD600. C) The calculated switching factor of JAy508 grown in different concentrations of neomycin.

An upwards trend was observed for both strains, JAy109 and JAy508, with the addition of increasing amounts of neomycin, suggesting a higher efficiency of repression with a greater amount of the ligand (Figure 4.15D) and 4.16C)). Comparison of the switching factors between JAy109 and JAy508 were carried out, comparing the activity of the riboswitch when expressed from a multi-copy 2μ vector versus a low copy vector (Re2.8) containing the CEN-ARS centromere.

The switching factors of JAy109 appear to be two times higher than that of strains containing the Re2.8 vector, suggesting that the riboswitch switching efficiency is higher in the pWHE601 (Figure 4.17A). The differences in switching factor can be due to multiple factors; plasmid copy number, reporter gene (GFP in pWHE601. YFP in Re2.8) or other elements on the plasmid such as the aptamer context and vector design (single vs. dual fluorescence vector). The aptamer context, in particular the

sequence immediately downstream of the M4 neomycin aptamer, can possibly affect the efficiency of the switching. The published pADH1-neo pWHE601 plasmid contains two in-frame start codons, with one within the aptamer sequence followed by a serine and lysine before the second methionine (Figure 4.17B – encircled with a green box). In contrast, the pADH1-neo Re2.8 plasmid contains three in-frame start codons. The first two, similar to the that found in the pADH1-neo pWHE601 (Figure 4.17B – encircled with a green box). However, this was followed by an additional valine, phenylalanine and the third methionine (Figure 4.17B – encircled with a red box). Differences in aptamer efficiency due to plasmid copy number and aptamer context have also been observed in other studies (Groher et al., 2018; Groher and Suess, 2014). Future work can investigate the effects of these sequences downstream of the aptamer.

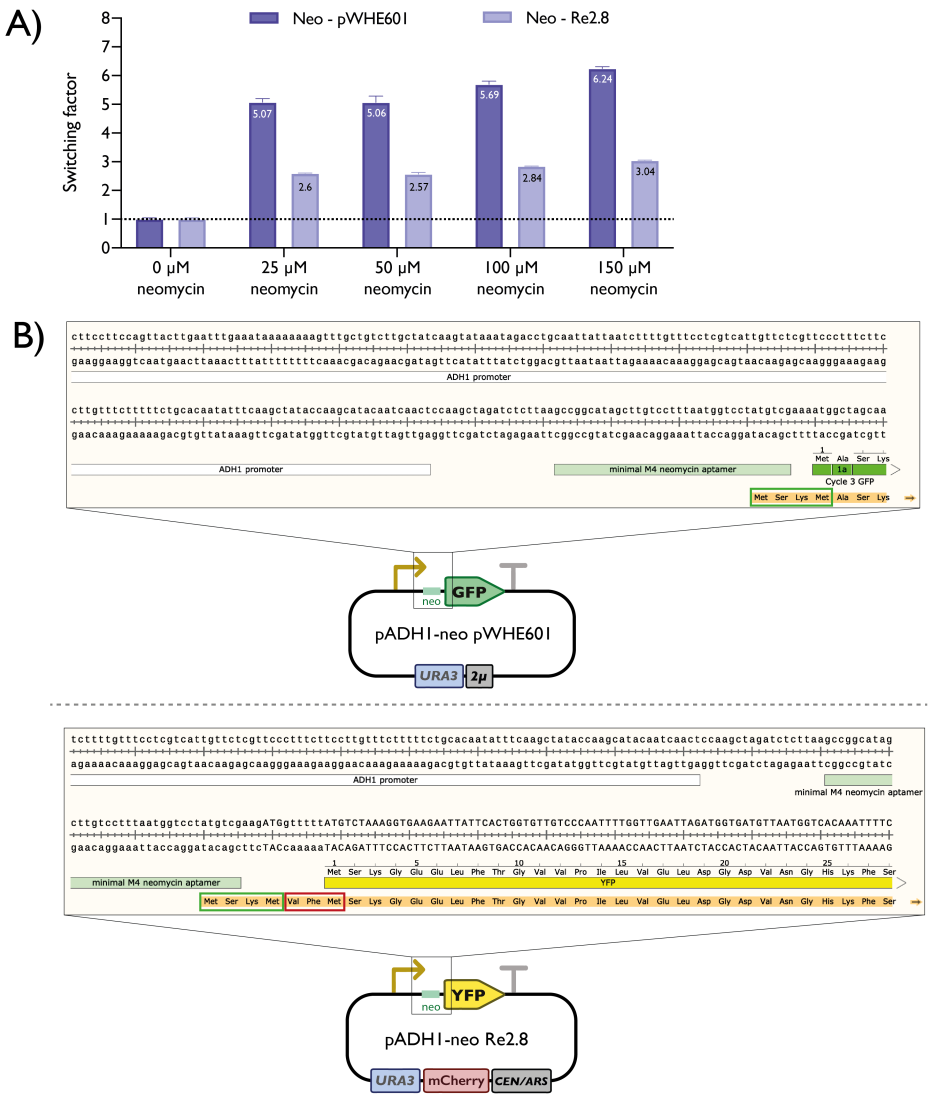


Figure 4.17

A) Comparison of switching factors of the neomycin riboswitch in low copy (CEN/ARS) vs. high copy (2μ) vectors. B) Differences in the aptamer and start codon context between the two plasmids. An additional Val, Phe and Met was introduced during the cloning process into the Re2.8 vector (labelled with a red box)

4.3 Chapter 4 discussion

Engineering of biosynthetic pathways necessitates optimisation of not only constituent genes but also their expression levels. One attractive option for modulation of

expression profiles is through the use of combinatorial libraries of endogenous derived promoter sequences.

It is however important to note that the length of the promoter sequence is arbitrary, with different publications utilising variable demarcations of a promoter sequence. A promoter sequence arbitrarily refers to sequence upstream of the start codon, encompassing the 5'UTR as well as several regulatory elements involved in controlling gene transcription. The arbitrary length defined is however crucial as it affects regions encompassed, and possible UAS sequences excluded, ultimately affecting the transcription rate, mRNA localisation and stability. As such when employing promoter sequences derived from the same endogenous promoter as existing studies, but using different constraints, activity should be screened again to reflect possible changes in promoter dynamics.

When assessing promoter dynamics over time, the promoter activities at exponential phase (OD 1) follow a highly similar trend in terms of promoter strength across all 74 promoters compared to the Guo et al data set (Guo et al., 2015). There are a few exceptions where the same level of expression was not recapitulated. Examples of this in the strong promoter library include S5-pOCH1 and S10-pRPL22C, where the measured promoter activity was lower than the weakest promoter in the library. This might be due to differences in the culturing and measurement method, resulting in a difference in promoter regulation. Further troubleshooting should be carried out to better understand the role that different measurement techniques, such as plate reader vs flow cytometry, play in observed expression results.

Promoter library characterisation has been limited to examining activities at one or a handful of time points, limiting our knowledge of the promoter functionality over time. As these promoters will be used for coordinating pathway expression for compound production over periods of 24 to 48 hours, knowledge of their activity profile over time is crucial. This is especially essential when the temporal control of enzyme production or the temporal limiting of a toxic enzyme or byproduct is required. In the work described in this chapter, I have utilised a plate reader to characterise the expression profiles of commonly used yeast promoters such as pTDH3, pPGK1, pACT1 and pCYC1. In particular, results from the characterisation of pPGK1 confirmed data previously obtained by Apel *et al.* at the time points they sampled (Apel et al., 2017). However, an increase in promoter activity at stationary phase ($\sim$10 hours) was not

previously seen, contradicting the notion of the activity of pPGK1 remaining constant, with minimal changes in expression over 24 hours of growth (Apel et al., 2017).

Further improvements on the assessing promoter dynamics could be achieved through the utilisation of destabilised fluorophores with a significantly reduced half-life. This would enable the changes in measurement of protein abundance (fluorescence) at any time point to much better reflect recent changed in transcription rate.

A disadvantage of native yeast promoters is that they are challenging to control, especially when constructing orthogonal pathways or circuits as they are tightly intertwined to the endogenous metabolism. Small molecule regulatable systems presents as a promising alternative, allowing greater control of protein production in a cell. In this chapter, I have investigated the activity of these regulatable systems (riboswitches and synthetic transcription factors) over 24 hours following induction. Characterisation of the neomycin riboswitch showed effective repression of gene expression following induction with at least 5 μM of neomycin. Growth curves have shown a decrease in growth with an increase in concentrations of neomycin introduced in the media. It would be useful to obtain the activity profile over time of the riboswitch with the addition of lower concentrations of neomycin (pM-nM concentrations).

Weigand and colleagues have tested the activity of the M4 neomycin aptamer inserted in the promoter of pTDH3 and pADH1 (Weigand et al., 2008). During this project, the M4 neomycin aptamer was inserted into several other yeast constitutive promoters (pPGK1 and pRPS3) and successfully verified by Sanger sequencing (Appendix Figure D.7. By incorporating the riboswitch driven by different strength promoters, it may be possible to have improved analysis of concentration across different concentration ranges, however due to time limitations, characterisation using the plate reader was not completed. Another regulatable promoter, pZ3EV was characterised in this chapter, showing upregulation of gene expression with inducer (estradiol) concentrations as low as 250 nM. The lowest concentration (250 nM) was adequate for the induction of gene expression to the same level as those for higher concentrations (2 μM). Apart from testing the Z_3EV system on the pGAL1 promoter (with inserted Zif268 binding motifs), McIsaac *et al.* have also tested the activity of pCYC1 containing these binding motifs resulting in an inducible promoter with a lower promoter activity (McIsaac et al., 2014).

In the characterisation of regulatable systems such as the riboswitches, a baseline fluorescence was detected despite saturating the culture with relatively high concentra-

tions of neomycin. It was difficult to determine whether the system had completely repressed the translation of the reporter or if a minimal fluorescence was still being produced. Future work will involve more accurate verification of the protein production, including flow cytometry and western blot to minimise or remove the impact of any background fluorescence. In addition, the screening of the strains would be preferably done in low fluorescence media in order to decrease the background fluorescence in the sample, however resulting changes in nutrient composition and associated impact upon metabolism may confound analysis and downstream applications.

Chapter 5

Constructing and diversifying the viomycin pathway

5.1 Chapter 5 introduction

5.1.1 A brief history of antibiotic discovery platforms

Historically, the driving force of drug discovery has been natural products and their chemically altered derivatives, with this group remaining an important source of antibiotics, constituting over 70% of antibiotics currently used in the clinic (Dashti et al., 2014). Between 1920 and 1950, a significant number of antibiotics from different microorganisms were discovered, resulting in this period being known as the "golden age of antibiotics" (Gould, 2016). Despite early successes, by the 1960s the returns from natural product-based drug discovery were diminishing due to high rediscovery of known compounds (Lewis, 2013). With a rise in antibiotic resistance, the pharmaceutical industry responded with a shift in their focus to synthetic antibiotics in part due to their ability to easily chemically produce small molecules, the development of high-throughput screening (HTS) methods and no requirement for culturing of challenging microorganisms (Li and Vederas, 2009; Thaker and Wright, 2015). Lipinski's Rule of Five was used as a guideline to drive the production of synthetic small molecule libraries, proposing that a drug has a higher probability of good absorption and permeation into a cell if the compound possesses certain qualities (less than 10 H+

bond acceptors and less than 5 H+ bond donors, calculated logP less than 5 and a molecular weight less than 500 Da) (Lipinski et al., 1997). However, Lipinski himself has pointed out that antibiotics violate the Rule of Five and in any case, screens of these libraries have shown limited success in generating antibiotic leads (Frantz, 2004; Wright, 2010). Although studies suggest a higher 'hit rate' for natural product screening versus synthetic compound libraries, difficulties still remain for the construction and screening of natural product libraries within reasonable time frames. As such new methodologies must be developed and employed before a paradigm shift can be expected in the pharmaceutical sector (Thaker and Wright, 2015)

5.1.2 Reinvigorating drug discovery: Using synthetic biology for drug discovery

In recent years developments in metabolic engineering and synthetic biology have facilitated the rapid refactoring and fine-tuning of natural product pathways. One famous example is the Semi-synthetic Artemisinin Project which utilised *S. cerevisiae* as a chassis for the mass production of artemisinic acid (25 g/L) (Paddon et al., 2013). *E. coli* was the initial chassis organism for heterologous expression of artemisinic acid due to simple and quick growth conditions, ease in carrying out genetic manipulations and the highly characterised metabolic and regulatory pathway. However, difficulties in expressing complex enzymes such as cytochrome p450, which was essential for generating the oxidised derivative of amorphadiene (artemisinic acid) have resulted in the shift to the more versatile *S. cerevisiae*. This combination of metabolic engineering and synthetic biology has resulted in the first success story in the industrial-level production of an essential drug for combatting malaria.

5.1.3 SynNP Project overview

One of the challenges for antibiotic drug discovery is the prohibitively high cost of research and development in natural products and the slow rate of novel drug discovery which, coupled with the rapid development of drug resistance, skew the cost-benefit ratio for biotechnological investment in this field. However, with the increasing focus on antimicrobial resistance over the past few years, greater incentives have been devised by governments and international organisations to encourage researchers and companies

to invest in reinvigorating the antibiotic pipeline (Organization, 2015). One current method employed for novel antibiotic drug discovery is through the use of tailoring enzyme. These are enzymes capable of chemically modifying a small molecule scaffold to provide or enhance therapeutic activity. Zhang et al. (2015) demonstrated this through the diversification of erythromycin using tailoring enzymes. Using this approach, they discovered novel variants effective against erythromycin-resistant *B. subtilis* (Zhang et al., 2015). Similarly, Yim et al. (2016) from the Wright laboratory have utilised the method for the generation of novel antibiotic compounds. They took a small group of glycopeptide antibiotic scaffold pathways and co-expressed these alongside a large library of tailoring enzymes in a *Streptomyces coelicolor* strain and successfully generated 9 novel molecules, some of which showed activity against glycopeptide resistant bacteria (Yim et al., 2016).

The 'SynNP' project, funded by the Gates foundation through the Global health program, was a collaborative research project with the Tyers laboratory at the University of Montreal and the Wright lab at McMaster University, looking to expand upon the approach demonstrated by Yim et al. (2016) and creating a high-throughput platform for natural product-based drug discovery that is applicable for any antibiotic class. As a proofs of principle, this project aimed to refactor (Figure 5.1A) and diversify several pathway scaffolds from different classes of natural products (Figure 5.1B) namely: viomycin (a non-ribosomal peptide), kirromycin (a polyketide), nocathiacin (a ribosomally synthesised and post-translationally modified peptide), capuramycin (a nucleoside) and naringenin (a flavonoid). The project was later expanded to incorporate Gram-positive antibiotic pathways such as rebeccamycin. The initial 5 exemplar pathways were chosen to represent the different classes of natural products that possesses antibiotic activity against *Mycobacterium tuberculosis*. According to the World Health Organisation (WHO), tuberculosis (TB) remains one of the top 10 causes of mortality worldwide with over 1.1 million deaths from the disease in 2015 (Organization, 2015). Currently, the treatment regimen for TB is a strict 6-month polytherapy course consisting of the 'intensive phase' with the regular consumption of four antibiotics (isoniazid, rifampicin, ethambutol and pyrazinamide) over 2 months and the continuation phase with isoniazid and rifampicin therapy for 4 months (Organization, 2015). However, this treatment regimen can often vary in resource-limited circumstances, which alongside poor adherence to regimens has contributed to the development of multidrug-resistant TB (MDR-TB). Patients infected with MDR-TB can be particularly challenging to treat due to the bacterial resistance to isoniazid and rifampicin, two front-line TB drugs (Organization, 2015). As a result, they will

require an extended treatment regimen with second-line TB drugs such as viomycin or capreomycin (intramuscualar injection), amikacin and fluoroquinolone with a poorer prognosis due to decreased potency and higher toxicity and side effects than front-line drugs. As such, there is a need for new drugs to tackle MDR-TB to help save millions of lives and accelerate the trend towards eradication. To address this, the project aims to establish a platform that would allow rapid production of synthetic-like natural products (SynNP) targeted against *M. tuberculosis* (Figure 5.1C).

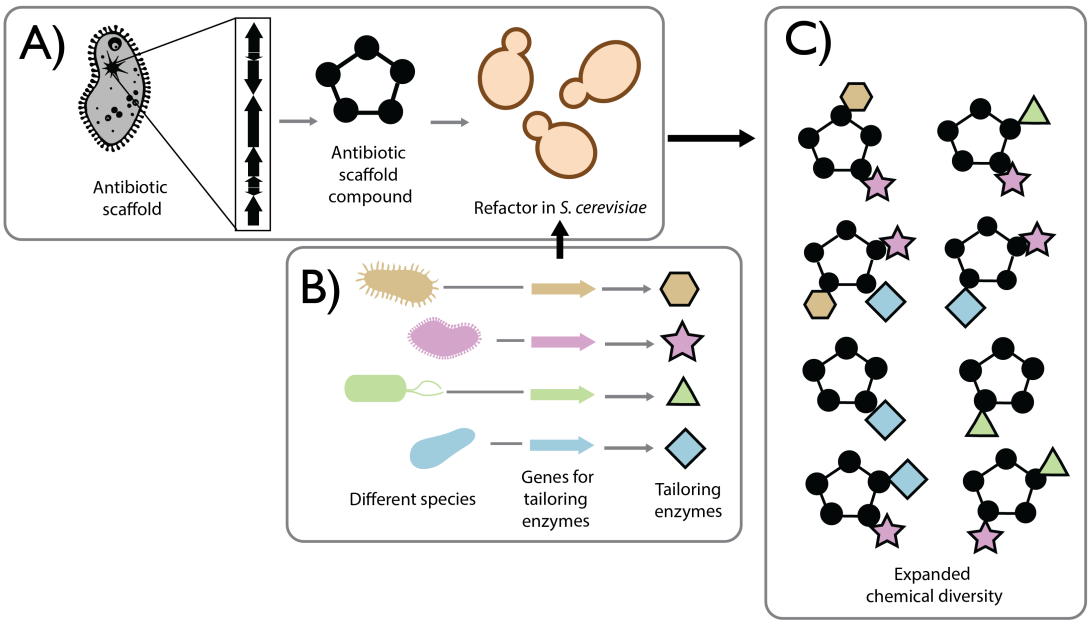


Figure 5.1

A) The refactoring of scaffold antibiotic biosynthetic pathways in yeast for the production of the antibiotic scaffolds (represented by the pentagon structure). B) The introduction of a library of tailoring enzymes (represented by the coloured arrows) from different microorganism into the yeast strain hosting the scaffold antibiotic biosynthetic pathway.C) The resulting library of novel antibiotics with different modifications (represented by the coloured shapes) introduced on the scaffold compound (pentagon structure). This represents an expanded chemical diversity of the scaffold antibiotic.

The main focus of this body of work was the refactoring of the viomycin pathway in yeast (Figure 5.1A). Upon successful production, the viomycin can then undergo subsequent modifications through the incorporation of tailoring enzymes, before ulti-

mately being screened for novel antibiotic function or improved activity (Figure 5.1B and C). The heterologous expression of biosynthetic pathways encoding viomycin was carried out using the YeastFab assembly method (Guo et al., 2015). Due to the limited number of transcription units (up to 6) permitted for assembly using this YeastFab toolkit, a colleague, Dr. Wei Liu expanded the YeastFab system to allow the assembly of up to 30 genes (unpublished data).

5.1.4 Viomycin

Viomycin is a non-ribosomal cyclic peptide and a member of the tuberactinomycin family of antibiotics, with activity against *M. tuberculosis*. It is commonly used as a second-line antibiotic to treat patients diagnosed with multi-drug resistant tuberculosis. Long term consumption of viomycin can result in side effects including ototoxicity, damaging both cochlear and vestibular function, as well as nephrotoxicity (Migliori and Zumla, 2017). As such the generation of altered variants providing wider therapeutic windows becomes highly desirable.

5.1.5 Viomycin mechanism of action

Viomycin affects the bacterial ribosomal machinery. Within a ribosome, three active sites are present (A site, P site and E site). The first tRNA (deacylated) enters in the P site and moves to the E site while the second aminoacyl tRNA enter through the A site and moves to the P site with the help of elongation factor G (EF-G), freeing up the A site for the third aminoacylated tRNA. Viomycin binds to the ribosome, at the interface between the large and small ribosomal subunits (Stanley et al., 2010). The viomycin interaction is dependent on the binding of a tRNA on the A site of the pre-translocation ribosome and is a competitive interaction between viomycin and EF-G. The successful binding of viomycin results in a conformational change that increases the tRNA affinity to the A site by 1,000-fold, resulting in the inhibition of translocation (Holm et al., 2016; Stanley et al., 2010).

5.1.6 Viomycin biosynthesis

Non-ribosomal peptides such as viomycin are synthesised in an assembly-line fashion by large multi-modular enzymes called non ribosomal peptide synthetases (NRPS) whereby each NRPS module sequentially incorporates an amino acid to a growing peptide product. As these peptides are synthesised independently of the ribosomal machinery, they are not limited to the 20 proteinogenic amino acids, contributing to their structural complexity. However, as each amino acid of the final peptide requires a specific NRPS module, gene clusters for non-ribosomal peptides tend to be large.

Viomycin has a cyclic pentapeptide core characteristic of members of the tuberactinomycin family of antibiotics. Each member of this family differs in their residue side chain modification (Stanley et al., 2010). Produced by *Streptomyces sp.* strain ATCC 11861, the viomycin pathway gene cluster consists of 21 biosynthetic enzyme genes (BSGs), *vioA* to *vioT* and *vph*. These enzymes are involved in the production of precursors, assembly of the viomycin cyclic pentapeptide core, side chain modifications, export and self-resistance (Thomas et al., 2003).

The NRPS involved in the synthesis of viomycin consists of VioA, VioF, VioI and VioG (Figure 5.2)(Barkei et al., 2009). Together, these enzymes assemble and cyclize two residues of L-2,3-diaminopropionate (L-Dap), two molecules of L-serine (L-Ser), and one molecule of L-capreomycidine (L-Cam) to form the minimal cyclic pentapeptide core (Barkei et al., 2009). L-capreomycidine is synthesised from L-arginine through the action of VioC and VioD. L-serine is the precursor for L-2,3-diaminopropionate and this conversion is catalysed by the combined actions of VioB and VioK (Thomas et al., 2003). Three subsequent modifications are introduced to this cyclic pentapeptide core to produce viomycin. The action of VioP along with the monomodular NRPS's, VioO and VioM catalyses the N-acylation of residue 1 with β-lysine (β-Lys). A further modification is introduced in residue 5, where VioQ catalyses the hydroxylation of L-capreomycidine in the cyclic ring. Finally, VioL catalyses the carbamoylation of residue 4 on the ring (Barkei et al., 2009) (Figure 5.2).

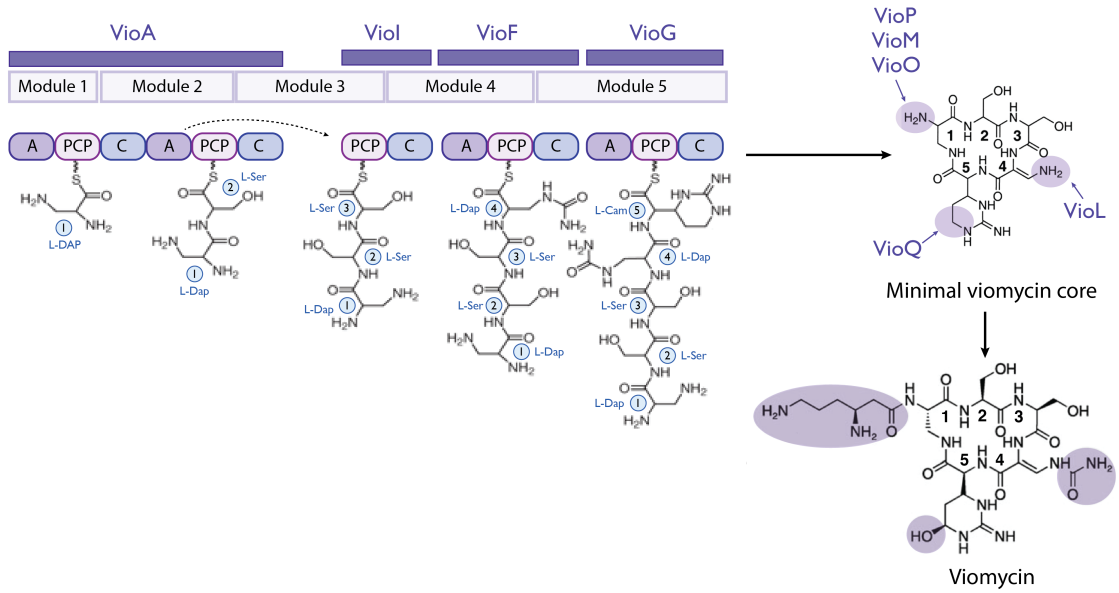


Figure 5.2

The production of the viomycin cyclic pentapeptide core using non-ribosomal peptide synthethases involving VioA, VioI, VioF and VioG. This pentapeptide core structure is further modified with the enzymes, VioP, VioM, VioO, VioQ and VioL to produce viomycin. Viomycin biosynthetic enzymes are in purple text.

In this body of work, I demonstrate the construction of a 20 BSG pathway in *S. cerevisiae* using the expanded YeastFab assembly method. The viomycin pathway is also combinatorially assembled with a library of promoters of different activities characterised in Chapter 4 to generate a library of strains with diverse expression profiles.

5.2 Chapter 5 results

5.2.1 Viomycin susceptibility tests)

The SynNP project relied on the refactoring of an antibiotic biosynthetic pathway in a different chassis organism. A crucial step prior to beginning the pathway assembly was ensuring that the antibiotic compound was not toxic to the chassis organism. Viomycin has not previously been shown to have any anti-fungal activity against *S.*

cerevisiae. In order to confirm this, different concentrations of viomycin (3-22 mM) were tested against wild type *S.cerevisiae* (BY4741) using a disk diffusion assay (Materials and Methods). No growth inhibition was observed in the presence of viomycin, across all concentrations tested (Appendix Figure E.1).

The SynNP platform also relied on high-throughput pathway assembly for the construction of large libraries of yeast producers. However, screening such these large libraries can be challenging. To mediate this, the SynNP pipeline relied on an initial antibiotic activity-based screening method against tester organisms. Within this project, two microorganisms (*M. smegmatis* and *B. subtilis*) were identified as relevant tester organisms and screened with different concentrations of viomycin using a disk diffusion assay. Growth of both microorganisms tested appeared to be inhibited in the presence of viomycin (Appendix Figure E.2).

5.2.2 Assembly of the ORF into the HcKan_O acceptor vector

The 19 viomycin BSGs used in this study were initially identified (*vioA*-vioS and*vph*), codon optimised for yeast expression and recoded to remove BsaI, BsmB1 and Sap1 restriction sites by Dr. Almer van der Sloot at the Tyers laboratory at the University of Montreal. Genes encoding transcriptional regulation (*vioR* and *vioT*) were excluded as they are likely non-functional when expressed in yeast. However, the subsequent pathway assembly methods varied between the collaborating groups, with one using Gibson assembly (Tyers lab) and another relying on Golden Gate cloning (Cai lab). As a result, the synthesis of the DNA was tailored to include 40 bp overhangs for Gibson assembly. These genes were then synthesised with these modifications incorporated, by a commercial DNA synthesis company (Gen9).

As these synthesised BSGs were not compatible with the YeastFab method for pathway assembly (which is the chosen method at the Cai lab), I designed primers to incorporate BsaI restrictions sites, the appropriate overhangs flanking every gene and a FLAG tag in the 3'end of the ORF. Viomycin BSGs were amplified to incorporate these elements by PCR using these specifically designed primers prior to their Golden Gate assembly to the parts-accepting vector (HcKan_O) (Figure 5.3A).

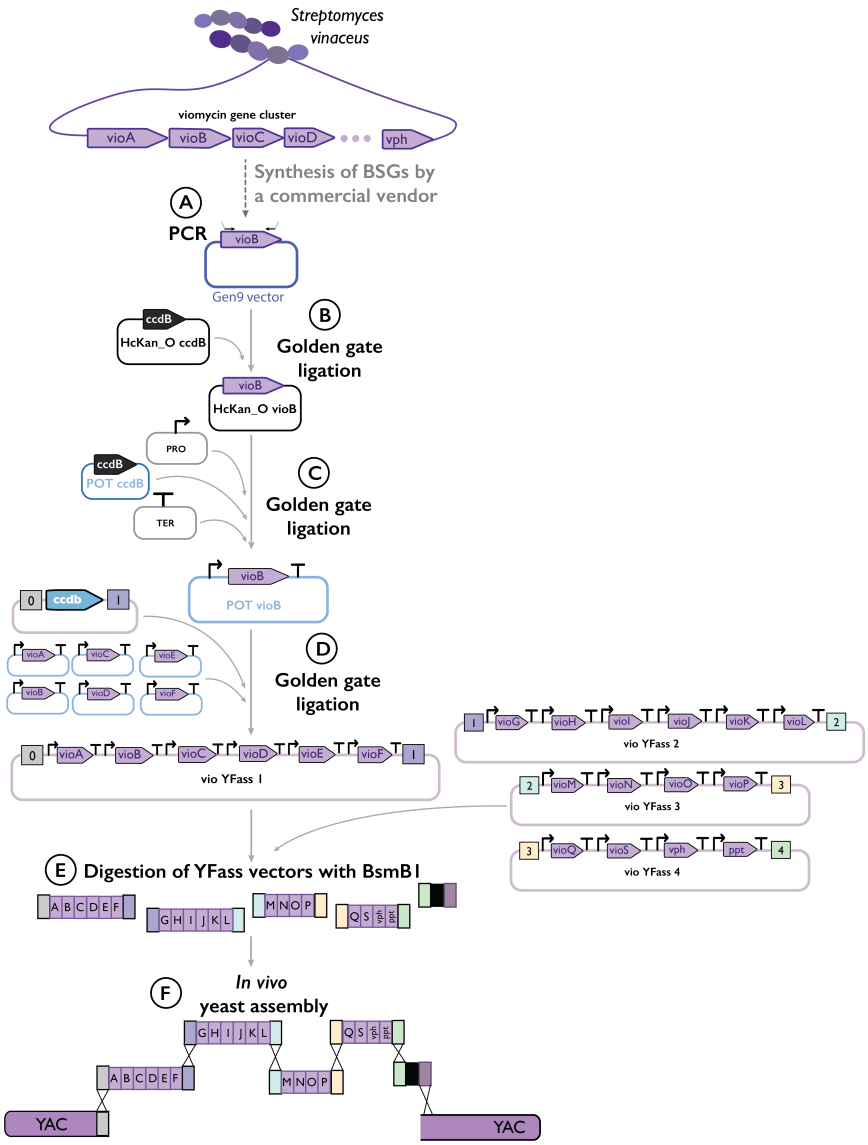


Figure 5.3

An overview of the viomycin assembly pipeline in the Cai lab. I began with A) the PCR amplification of each BSG from the previously synthesised BSGs to incorporate the BsaI restriction sites and appropriate 4 bp overhang. B) These PCR products were used for the domestication into the HcKan_O parts accepting vector by golden gate cloning. C) Assembly of a transcription unit, including a promoter, BSG and a terminator into a POT parts accepting vector is carried out by golden gate cloning. D) These transcription units can be assembled together (maximum of 6) in a multi-transcription unit vector (YFass) by another round of golden gate cloning. E) The joined multi-transcription units are released from their respective YFass vectors by digestion with BsmBI, revealing homology arms. F) These homology arms aid in the assembly of the full 20 gene pathway in a linear YAC by *in vivo* yeast homologous recombination. All of this work was done by myself at the Cai lab.

As the YeastFab assembly method relied on hierarchical Golden Gate cloning, each of the 20 genes needed to be domesticated into the YeastFab format. Primers were ordered from a commercial vendor (IDT) and were specifically designed to amplify each gene and introduce the BsaI restriction site and the GATG prefix and TAGC suffix flanking each BSG (Table of primers). The previously synthesised viomycin genes were used as a template for the PCR amplification (Figure 5.3A) as described in the Materials and Methods section and the resulting amplicon was subsequently used as the insert for the Golden Gate assembly into our ORF parts-accepting vector (HcKan_O) (Figure 5.3B). Upon cloning, the plasmids were initially screened by colony PCR and promising clones were verified by Sanger sequencing using the HcKan chc forward and reverse primers (Table of primers). All 20 viomycin BSGs were successfully cloned into the HcKan_O vector with no mutations introduced (Appendix Figure E.3).

5.2.3 Assembly of the full pathway with a set of 'strong promoters'

The full viomycin pathway was assembled using the expanded YeastFab cloning method, allowing the assembly of up to 30 genes. Using the domesticated viomycin genes in the previous section, transcription units consisting of a promoter, ORF and a terminator were assembled into POT vectors by Golden Gate cloning (Materials and Methods section) (Figure 5.3B). Promoters, terminators and POT acceptor vectors for the assembly of each viomycin BSG are outlined in (Table 5.1), different POT vectors were utilised too enable downstream cloning of the full pathways. Promoters used in this assembly were from the 'strong' promoter collection characterised in Chapter 4. Following the golden gate ligation, each assembly was digested to verify the size of the insert (Table 5.1). The digestion patterns for all 20 viomycin BSG POTs (Figure 5.4A) show the same pattern as seen in the simulation (Figure 5.4B).

Table 5.1 Promoters, terminators and POT vectors used for each viomycin BSG assembly (strong promoter assembly)

POT name	Promoter	ORF	Terminator	POT vector	Total size of POT insert (bp)
*vioA*_POT2	pTDH3	*vioA*	tACS2	POT2	7238
*vioB*_POT4	pENO2	*vioB*	tMDM10	POT4	1704
*vioC*_POT6	pHXK1	*vioC*	tTPD3	POT6	1749
*vioD*_POT8	pHSP12	*vioD*	tPMT2	POT8	1925
*vioE*_POT10	pHSP12	*vioE*	tCIT1	POT10	2099
*vioF*_POT11	pZEO1	*vioF*	tFUM1	POT11	3885
*vioG*_POT2	pSTE2	*vioG*	tCDC24	POT2	4019
*vioH*_POT4	pRPL27A	*vioH*	tSPC72	POT4	1544
*vioI*_POT6	pRPS27B	*vioI*	YAL056C-A_T	POT6	2408
*vioJ*_POT8	pACT1	*vioJ*	YAL063C_T	POT8	1928
*vioK*_POT10	pFUR1	*vioK*	YAL067W-A_T	POT10	1838
*vioL*_POT11	pPNC1	*vioL*	tSEN34	POT11	1682
*vioM*_POT2	pPGK1	*vioM*	tENO2	POT2	2129
*vioN*_POT4	pCRH1	*vioN*	tSWC3	POT4	855
*vioO*_POT6	pRTC3	*vioO*	tPSK1	POT6	2588
*vioP*_POT7	pACT1	*vioP*	tACS1	POT7	2093
*vioQ*_POT2	YHR097C_P	*vioQ*	tGIP4	POT2	1732
*vioS*_POT4	YGL072C_P	*vioS*	tRBG1	POT4	2114
*vph*_POT6	pAGA2	*vph*	tRFA1	POT6	1637
transferase_ POT7	pHAC1	*ppt*ase	tSEO1	POT7	1427

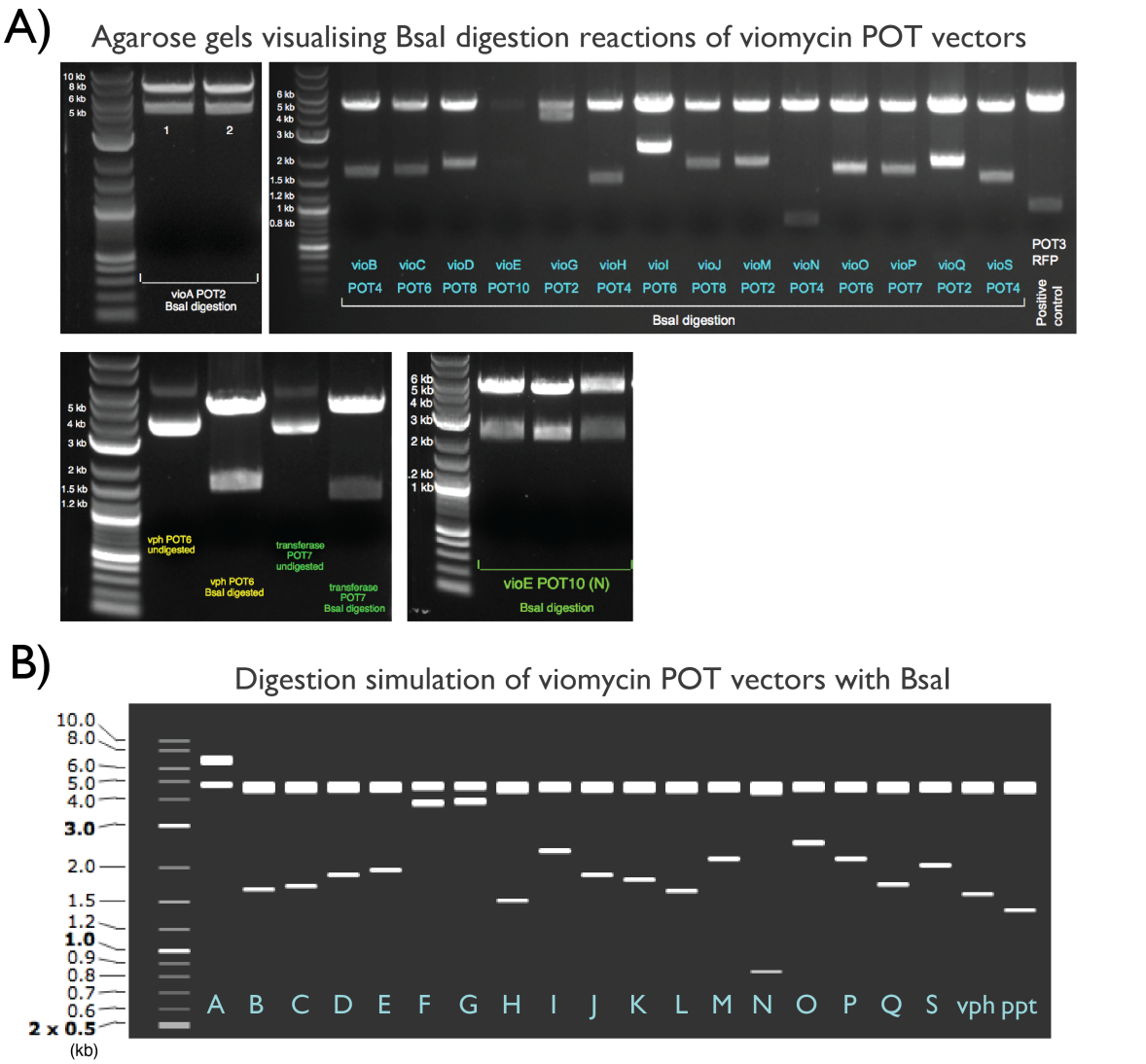


Figure 5.4

A) Agarose gels visualising the digestion reaction for all 20 viomycin POTs following BsaI digestion. B) Digestion simulation of all 20 viomycin POTs (using the software SnapGene software) to illustrate the expected band pattern.

The assembly of multiple transcription units were assembled into YFass vectors based on the assembly plan outlined in Table 5.2, using Golden Gate cloning (Figure 5.3D). Rows are coloured based on which POTs were to be assembled together into a YFass vector. A fifth empty YFass vector was previously constructed, only containing the homology arms.YFass vectors were screened by restriction digestion to verify the size of the assembly in each YFass vector.

Table 5.2 Plan for the multi-transcription unit assembly for the viomycin pathway (Strong promoter assembly)

YFass name	POT	YFass vector	Total size of multi-TU insert (bp)
vio_YFass 1	*vioA*_POT2	YFass 1	18600
	*vioB*_POT4		
	*vioC*_POT6		
	*vioD*_POT8		
	*vioE*_POT10		
	*vioF*_POT11		
vio_YFass 2	*vioG*_POT2	YFass 2	13419
	*vioH*_POT4		
	*vioI*_POT6		
	*vioJ*_POT8		
	*vioK*_POT10		
	*vioL*_POT11		
vio_YFass 3	*vioM*_POT2	YFass 3	7665
	*vioN*_POT4		
	*vioO*_POT6		
	*vioP*_POT7		
vio_YFass 4	*vioQ*_POT2	YFass 4	6910
	*vioS*_POT4		
	*vph*_POT6		
	transferase_POT7		

The digestion patterns for all 4 viomycin YFass (hereafter vio YFass) vectors show the same pattern as seen in the simulation (Figure 5.5A and B). These vio YFass vectors were also verified for the presence of each viomycin gene by PCR, using the primers used to amplify and incorporate BsaI sites flanking each gene (Table of primers) (Figure 5.5C). All 20 genes were present in all 4 vio YFass vectors (Figure 5.5D). Successfully verified assemblies were digested with BsmBI to release each of the multi-transcription units with their respective homology arms from the YFass backbone vector (Figure 5.3E). The pYAC10 pathway acceptor vector was digested with SacI and XbaI to linearise the YAC, revealing the homology arms. The pYAC10

was also digested with BamHI to fragment the YAC into two, producing a left arm and a right arm. The left arm of the YAC possessed a HIS auxotrophic marker and the right YAC arm possessed a LEU marker. All 5 YFass vectors were assembled to a linear YAC in vivo by yeast homologous recombination and screened on SC -HIS -LEU plates (Figure 5.3F).

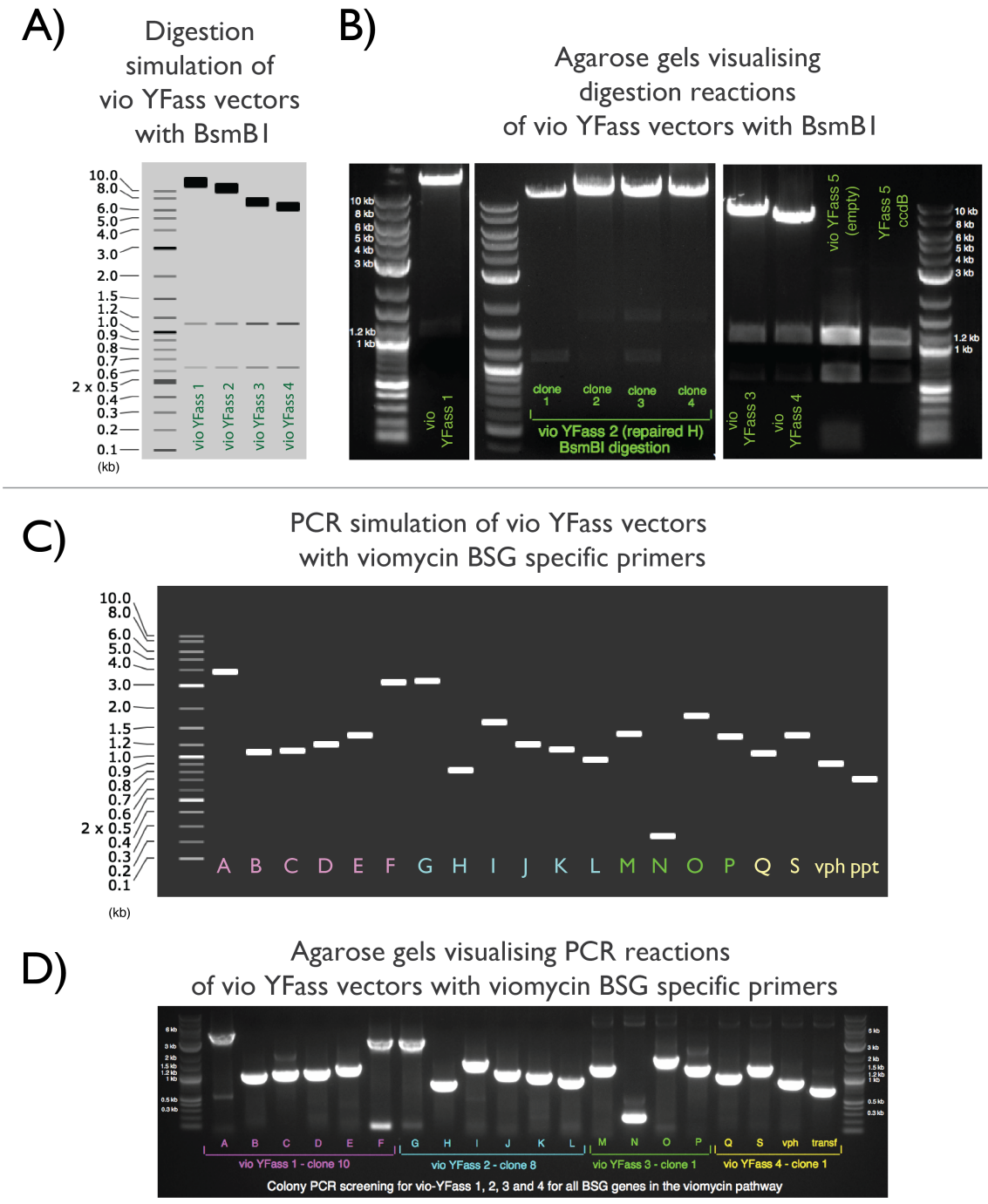


Figure 5.5

A) Digestion simulation for all 4 viomycin YFass assemblies. B) Agarose gel showing the digestion verification results. C) PCR simulation of each gene in all 4 viomycin YFass vectors. D) Agarose gel verifying the PCR verification of all 20 genes.

24 clones (vioYAC1-vioYAC24) were isolated from the transformation plate and screened for the presence of 4 genes (*vioC*,*vioJ*,*vioO* and *vph*), each derived from a different original YFass vector (Figure 5.6A). The yeast colony PCR was carried out according to the protocol described in Materials and Methods using gene specific primers (Table of primers). A band for *vioC* was observed in 22 out of 24 clones screened, 20 out of 24 clones showed an amplicon for *vioJ*, 23 out of 24 clones showed an amplicon for *vioO* and all clones showed a band for *vph*. Clones that showed missing bands were vioYAC 7, 12, 13 and 15, with the last three missing two genes (Figure 5.6B). This suggests an ∼80% success rate in a full pathway assembly of 20 genes. One viomycin YAC clone (vioYAC1 – JAy001) was selected and screened by colony PCR for the presence of all 20 genes in the viomycin pathway using primers targeting a 200 bp region of the ORF (Table of primers) (Figure 5.6C). Based on the colony PCR results generated (Figure 5.6D), all the genes appear to be present; however one band did appear in the negative control suggesting the presence of binding sites for this primer pair within the yeast genome. Importantly, the band size was larger than 200 bp. The size difference provided sufficient motivation to proceed with the sequence verification. Verification of expression of each viomycin BSG gene by RT-PCR was carried out. The agarose gel visualising the RT-PCR products of JAy001 show a band of the correct size (200 bp) for all 20 genes (Figure 5.6E). JAy001 was sequenced by Beijing Genomics Institute (BGI) and the data was processed by Marco Monti and Dr. Daniel Schindler. Alignment of the reads to the JAy001 reference in IGV shows full coverage of all the viomycin genes on the YAC, with no reads observed for the URA3 selection cassette flanked by telomeres. This was because the BamHI digested the URA3 cassette, linearising the YAC and exposing the telomeres at either end (Appendix Figure E.4)

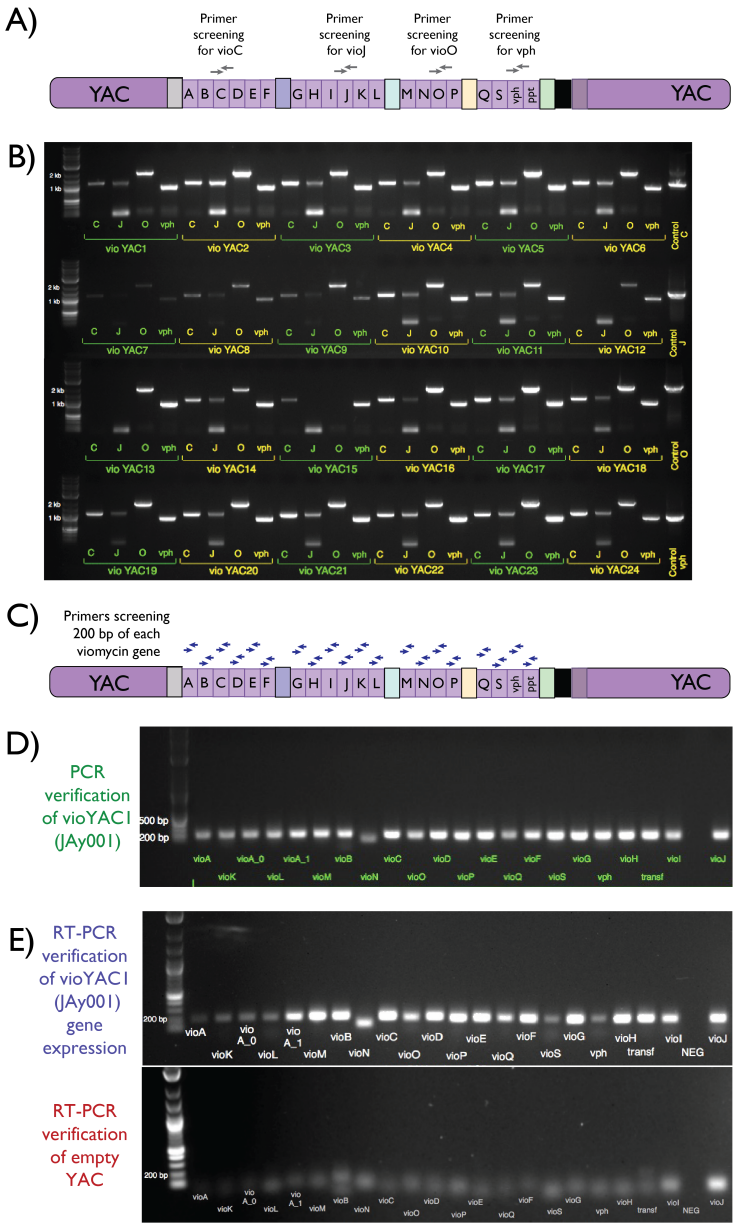


Figure 5.6

A) A diagrammatic representation of the final viomycin pathway assembly and the location of the 4 primer pairs used for the initial yeast colony screening. B) An agarose gel illustrating the colony PCR amplification results for 24 yeast strains containing the viomycin pathway C) A diagrammatic representation of the viomycin pathway and the location of the primers for the full pathway screening D) An agarose gel visualising the colony PCR verification of all 20 genes in the vioYAC1 (JAy001) strain. E) RT-PCR verifying the gene expression of all 20 BSGs in the JAy001 contrasted to a negative control (empty YAC)

.A preliminary cell-based screening assay was used to determine whether JAy001 was producing a bioactive compound. This cell-based assay relied on the growth of the yeast producer strain on an agar plate, which is subsequently overlaid with phosphate buffered agar inoculated with a tester strain (*Bacillus subtilis*, *Micrococcus luteus* or *Mycobacterium smegmatis*) (Figure 5.7A). Yeast strains producing an antibiotic compound should show a zone of inhibition or 'halo' surrounding the yeast strain (Figure 5.7B). In this case, the JAy001 strain was grown on an SC -HIS -URA plate for 48 hours. A layer of molten agar inoculated with *B. subtilis* was added onto the JAy001 yeast plate and grown for a subsequent 24 hours. The cell-based screening of JAy001 showed no zones of inhibition surrounding the yeast colonies, suggesting no production of an antibiotic compound active against *B. subtilis* (Figure 5.7C). A violacein-producing strain was used as a positive control for the cell-based assay, showing the presence of zones of inhibition surrounding the yeast colonies. Subsequent validation of JAy001 spent media and cell pellets by mass spectrometry showed no production of viomycin within detectable limits.

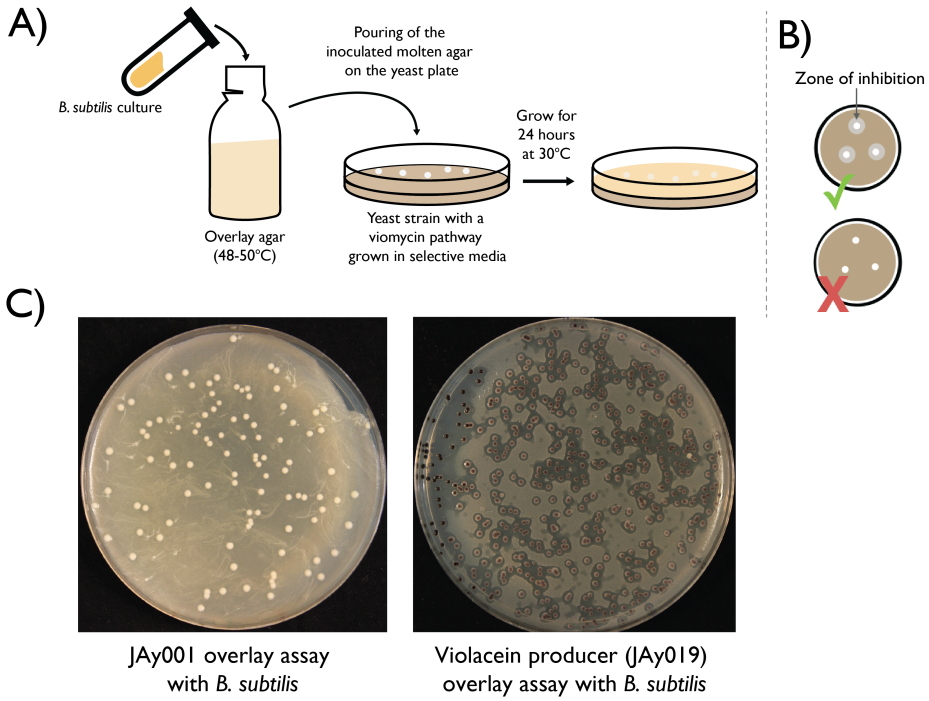


Figure 5.7

A) A diagram illustrating the overlay assay used to probe for the production of a bioactive compound against a tester strain (*B. subtilis*). In this assay, *B. subtilis* is added to molten overlay agar (48-50°C) then added on top of a yeast agar plate previously grown with a yeast strain containing the viomycin pathway. The overlay agar is allowed to set before incubating it for 24 hours at 30°C. B) Yeast strains producing a bioactive product will show a zone of inhibition or 'halo' surrounding the yeast colony C) The results for the JAy001 overlay assay with *B. subtilis*. The positive control for this assay is the growth of *B. subtilis* in violacein producing yeast strains.

5.2.4 Combinatorial assembly of viomycin pathway scaffolds with different promoters

The alteration of expression for each gene in the pathway was carried out through the combinatorial assembly of viomycin BSGs with a pool of promoters of different strengths. To prevent repeated use of the same promoter and decrease the risk of rearrangements due to unwanted recombination, unique promoter sub-libraries were chosen for each transcription unit (Figure 5.8A)(Table 5.3). Each sub-library contains three promoters of strong, moderate and weak strength as characterised in Chapter 4.

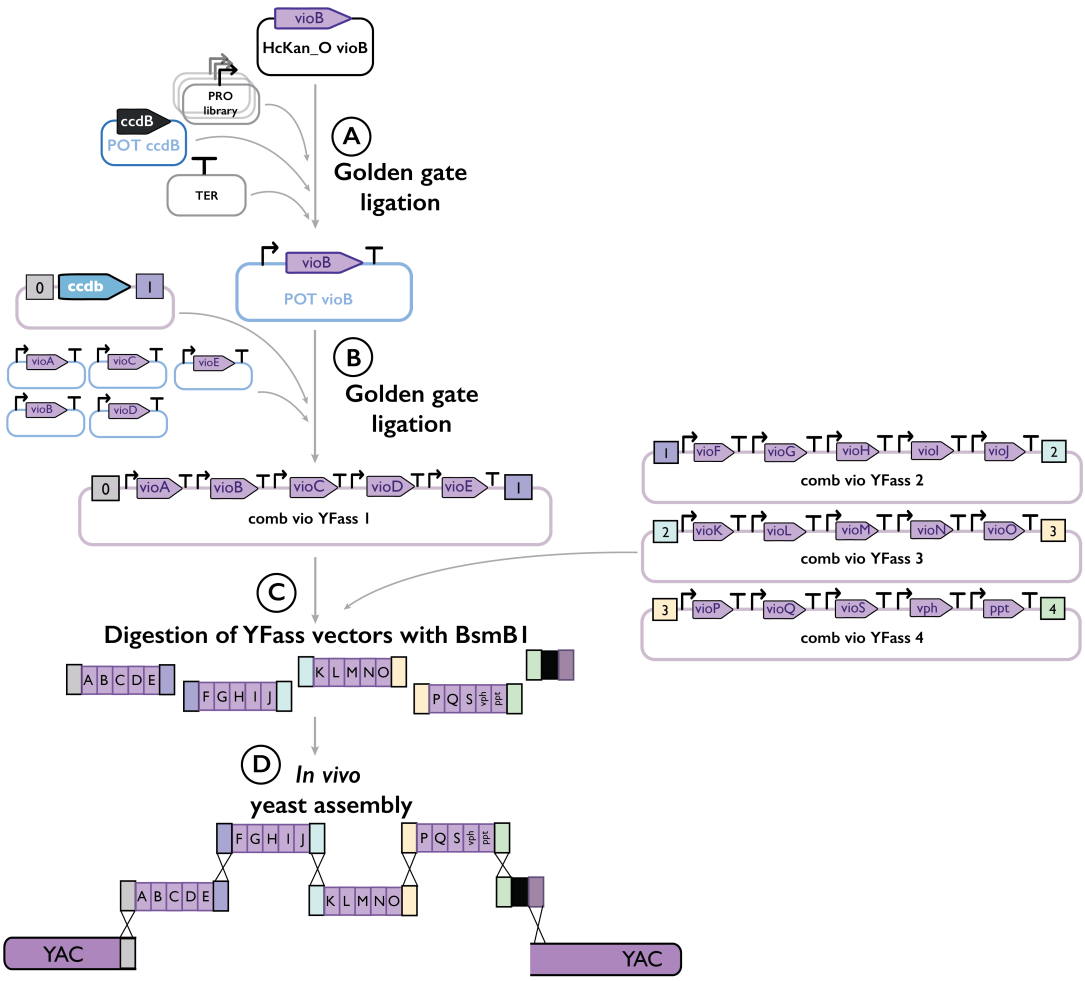


Figure 5.8

Combinatorial assembly viomycin BSGs with a pool of promoters of different activities. Beginning with A) Combinatorial assembly of a transcription unit, with a pool of promoters, a BSG gene and a terminator into a POT parts accepting vector by golden gate cloning. B) These transcription units can be assembled together (maximum of 6) in a multi-transcription unit vector (YFass) by another round of golden gate cloning. C) The joined multi-transcription units are released from their respective YFass vectors by digestion with BsmBI, revealing homology arms. D) These homology arms aid in the assembly of the full 20 gene pathway in a linear YAC by in vivo yeast homologous recombination.

Table 5.3 Promoters, terminators and POT vectors used for each viomycin BSG assembly. Each BSG was assembled with a pool of 3 promoters in a combinatorial assembly.

Promoters in the pooled library			ORF	Terminator	POT vector	Final vector name	Insert size (bp)
pTDH3	pAAD6	pYGR237C_P	*vioA*	tACS2	POT2	comb_ *vioA*_POT2	7238
pENO2	pDIA4	pOKP1	*vioB*	tMDM10	POT4	comb_ *vioB*_POT4	1704
pHXK1	pGPA1	pYPI1	*vioC*	tTPD3	POT6	comb_ *vioC*_POT6	1749
pHSP12	pYPT1	pSMX2	*vioD*	tPMT2	POT8	comb_ *vioD*_POT8	1925
pOCH1	pSBP1	pPRE4	*vioE*	tCIT1	POT9	comb_ *vioE*_POT9	2099
pSTE2	pBCD1	pHSN1	*vioF*	tFUM1	POT2	comb_vioF_POT2	3885
pRPS27B	pLSC2	pMRP4	*vioG*	tCDC24	POT4	comb_vioG_POT4	4019
pACT1	pGRE3	YHL005C_P	*vioH*	tSPC72	POT6	comb_vioH_POT6	1544
pPNC1	pSBE22	YGR161W-C_P	*vioI*	YAL056C-A_T	POT8	comb_vioI_POT8	2408
pRPL22B	pFSH1	pSET5	*vioJ*	YAL063C_T	POT9	comb_vioJ_POT9	1928
pCRH1	pRPL2A	pLSM12	*vioK*	YAL067W-A_T	POT2	comb_vioK_POT2	1838
YHR097C_P	pELP2	pRNA15	*vioL*	tSEN34	POT4	comb_vioL_POT4	1682
pSOD2	pVMA10	pRBG2	*vioM*	tENO2	POT6	comb_vioM_POT6	2129
pAGA2	pPAU13	pOPI1	*vioN*	tSWC3	POT8	comb_vioN_POT8	855
pHAC1	pRPN1	pGTR2	*vioO*	tPSK1	POT9	comb_vioO_POT9	2588
pPXR1	pUBP6	pRIM15	*vioP*	tACS1	POT2	comb_vioP_POT2	2093
pSCW4	pSSZ1	pGAR1	*vioQ*	tGIP4	POT4	comb_vioQ_POT4	1732
pQCR9	pSLT2	pDST1	*vioS*	tRBG1	POT6	comb_vioS_POT6	2114
YHR162W_P	pERG1	pCGR1	*vph*	tRFA1	POT8	comb_vph_POT8	1637
pRIM101	pNIC96	pTFG1	*ppt*ase	tSEO1	POT9	comb_pptase_POT9	1427

Each of the 20 BSG were combinatorially assembled with a promoter pool and a terminator into POT vectors (Table 5.3) by Golden Gate assembly using Esp3I. This assembly is illustrated in (Figure 5.8A) except that instead of a defined promoter, a pool of promoters was used instead. Products of the cloning reaction were isolated by pooling all the colonies from the transformation plate instead of picking single colonies. in order to capture diverse combinations from the POT assembly reactions. Digestion verification with BsaI restriction enzyme showed the release of viomycin transcription units of the correct size from the mixed construct pools of POT vectors (Figure 5.9).

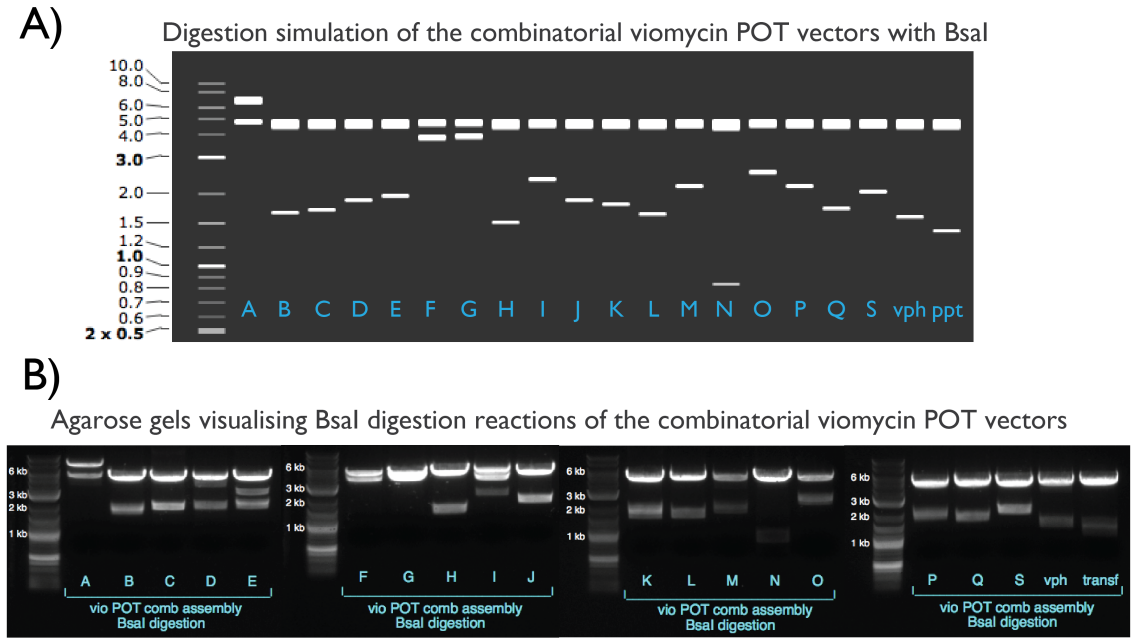


Figure 5.9

Digestion simulation and agarose gel verification of all 20 combinatorial viomycin POT assemblies

The pooled viomycin POTs were used for the subsequent multi-transcription unit assemblies into the YFass vectors by adding the appropriate POT vectors together in 4 separate Golden Gate assembly reactions as outlined in Table 5.4 and transformed into *E. coli* cells (Figure 5.8B). Similar to the previous combinatorial viomycin POT assembly, the products of the combinatorial assembly were isolated by pooling all the colonies from the transformation plate. The combinatorial vio YFass vectors were digested with Esp3I to verify the size of the inserts into the multi-TU vector (Figure 5.10A for the digestion simulation and B for the gel result). Although the expected band sizes were present in the agarose gel verification of the digestion reaction, other

smaller bands were also observed suggesting some misassemblies (Figure 5.10). Future work will include the optimisation of this assembly step to decrease the number of misassembles in pooled reactions.

Table 5.4 Assembly plan for the multi-transcription unit assembly for the combinatorial viomycin pathway and the calculated size of the final multi-transcription unit insert (bp)

YFass name	POT	YFass vector	Total size of multi-TU insert (bp)
comb_vio_YFass 1	comb_ *vioA*_POT2	YFass 1	14715
	comb_ *vioB*_POT4		
	comb_ *vioC*_POT6		
	comb_ *vioD*_POT8		
	comb_ *vioE*_POT9		
comb_vio_YFass 2	comb_vioF_POT2	YFass 2	13784
	comb_vioG_POT4		
	comb_vioH_POT6		
	comb_vioI_POT8		
	comb_vioJ_POT9		
comb_vio_YFass 3	comb_vioK_POT2	YFass 3	9092
	comb_vioL_POT4		
	comb_vioM_POT6		
	comb_vioN_POT8		
	comb_vioO_POT9		
comb_vio_YFass 4	comb_vioP_POT2	YFass 4	9003
	comb_vioQ_POT4		
	comb_vioS_POT6		
	comb_vph_POT8		
	comb_pptase_POT9		

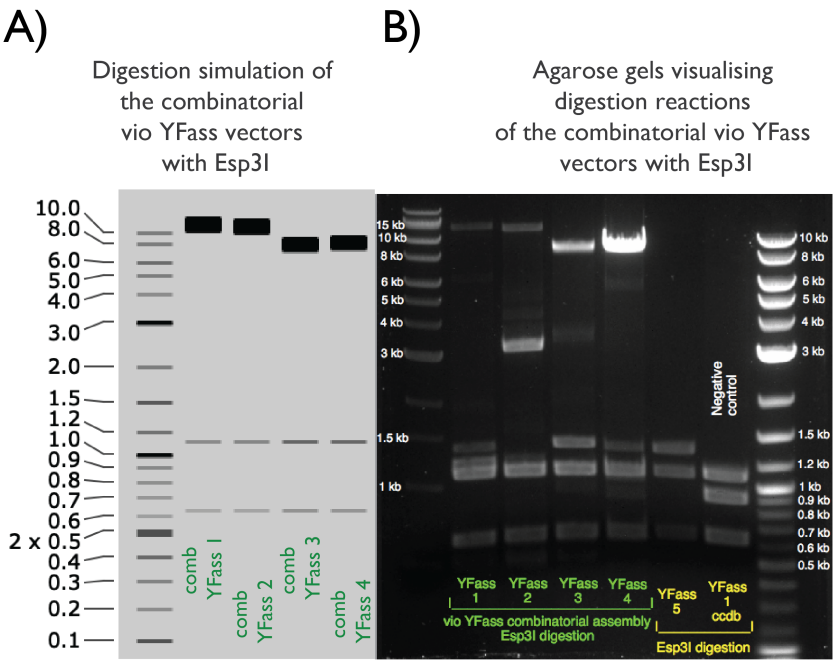


Figure 5.10

Digestion simulation and agarose gel verification of all 4 combinatorial viomycin assemblies in the YFass vector

To determine the current efficiency of the assembly pipeline, these Esp3I digested combinatorial vio YFass vectors were transformed along with the previously digested pYAC10 into yeast for the subsequent assembly of the viomycin pathway into a linear YAC by homologous recombination (Figure 5.10D). As the left arm of the YAC possessed a HIS auxotrophic marker and the right YAC arm possessed a LEU marker, the yeast transformation was selected on SC -HIS -LEU agar plates. 96 YAC strains were picked and screened for the presence of the full pathway by colony PCR.

The first screening step involved PCR verification using 4 pairs of primers, each targeting a single viomycin BSG from each of the 4 YFass segments (*vioE*,*vioJ*,*vioO*,*vph*) (Figure 5.11A). 50% of clones screened resulted in the detection of all 4 genes by PCR verification (Figure 5.11B and Appendix figure E.5). These 48 clones were subsequently screened for the presence of all 20 genes in the viomycin pathway (Figure 5.11C) and 92% showed an amplicon for the complete set of 20 genes. Only 4 clones showed 1 missing amplicon (*vioK*). Overall, the current assembly method for large pathways showed a success rate of 46% (Figure 5.11D and Appendix figure E.6).

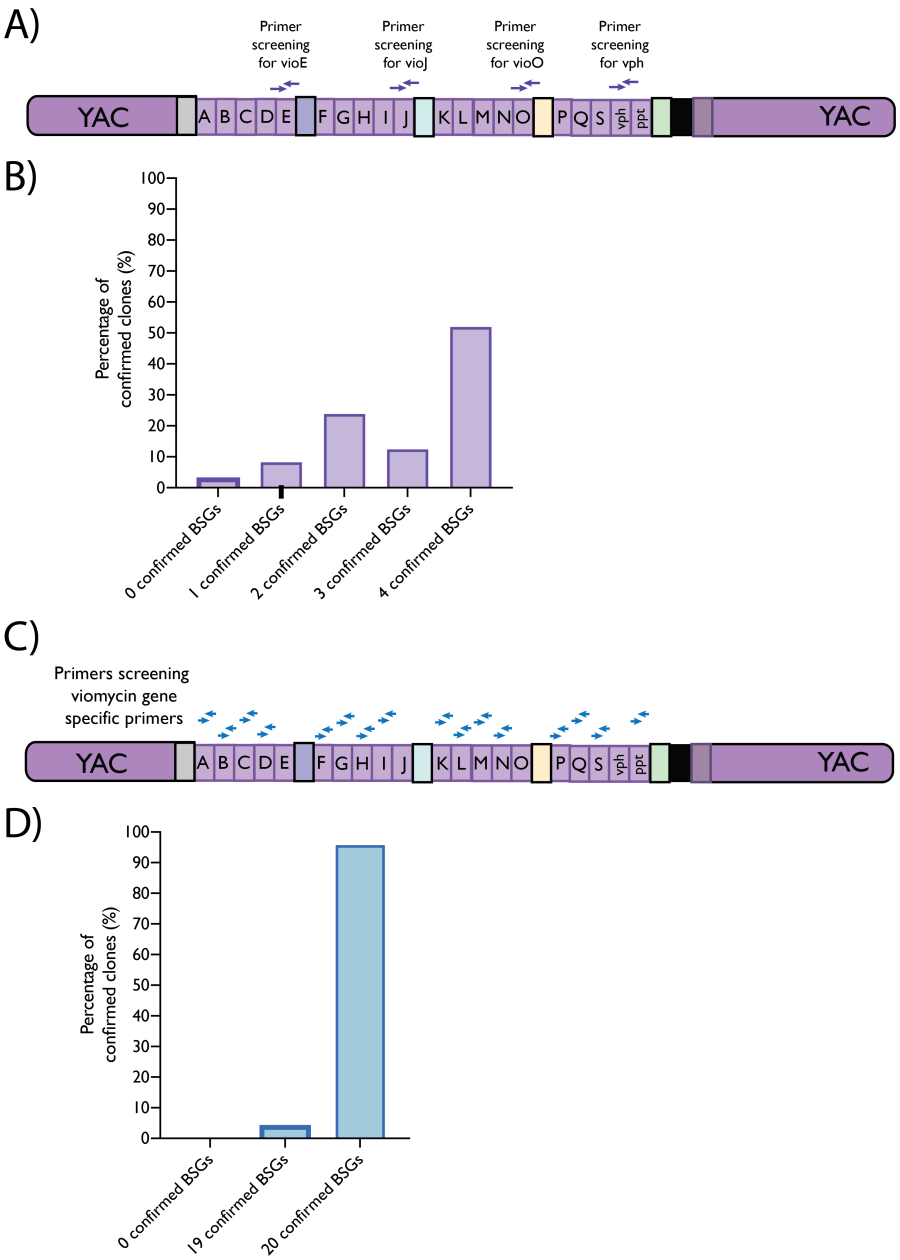


Figure 5.11

A) Diagrammatic representation of the primers used for the initial colony PCR screening of 96 clones from the combinatorial assembly of the viomycin pathway using primers targeting 4 genes (vioE,*vioJ*,*vioO*,*vph*). B) The percentage assembly success rate based on initial screening of 96 clones using 4 pairs of primers. Gel images for these colony PCR verification experiments are shown in Appendix figure E.5. C) Diagrammatic representation of the primers used for the full pathway screening for the 48 clones containing 4 confirmed BSGs. D) The percentage success rate based on initial screening of 48 clones using 20 pairs of primers. Gel images for these colony PCR verification experiments are shown in Appendix figure E.6

All 96 clones from the combinatorial viomycin assembly were screened using the overlay assay. For this large-scale screening, the overlay assay was modified such that all 96 clones were first grown in liquid culture for 24 hours. A small amount (5 μL) of the culture is used to spot each clone on an omnitray (Thermo fisher) containing yeast selective agar and grown for 48 hours prior to the addition of overlay agar inoculated with *B. subtilis* (Figure 5.12A). Clones producing an antibiotic compound are those showing a 'halo' or zone of inhibition surrounding the yeast colony (Figure 5.12B). All 96 clones were grown in an SC-HIS -LEU plate (Figure 5.12C) and SC -HIS -LEU dropout plate supplemented with precursors for the viomycin pathway (ornithine, lysine, arginine and serine) (Figure 5.12D); however, three clones appeared to have not grown, perhaps due to the loss of the YAC. None of the clones in either plate appeared to produce compounds that inhibit *B. subtilis* growth (Figure 5.12C and D). This suggests that either the yeast strains do not produce viomycin or any other antibiotic compound, or a low amount below the minimum inhibitory concentration (MIC) is produced resulting in no visible inhibition of bacterial growth. Six random clones, comb vioYAC 4, comb vioYAC 24, comb vioYAC 35, comb vioYAC 36, comb vioYAC 48 and comb vioYAC 95 were chosen for further verification by LC-MS. This extraction and chemical analysis were carried out by Dr. Venkateswarlu Yarlagadda from the Wright lab at McMaster University according to the protocol described in Materials and Methods section - Viomycin extraction and screening by LC-MS for the production of viomycin and its intermediates. However, there was no evidence of viomycin production or its intermediates in any of the strains screened by LC-MS. In addition, Dr. Grace Yim screened these extracted fractions by disk diffusion assay against *B. subtilis* and found no antibacterial activity in the fractions tested (Appendix Figure E.7). As none of the variants in the combinatorial assembly library showed the production of viomycin or its intermediates, I proceeded to truncate the pathway to identify non-functional enzymes or bottlenecks in the pathway. This truncated pathway would only consist of enzymes for the production of the viomycin cyclic pentapeptide pathway, and excluding enzymes for the side chain modification of the core structure.

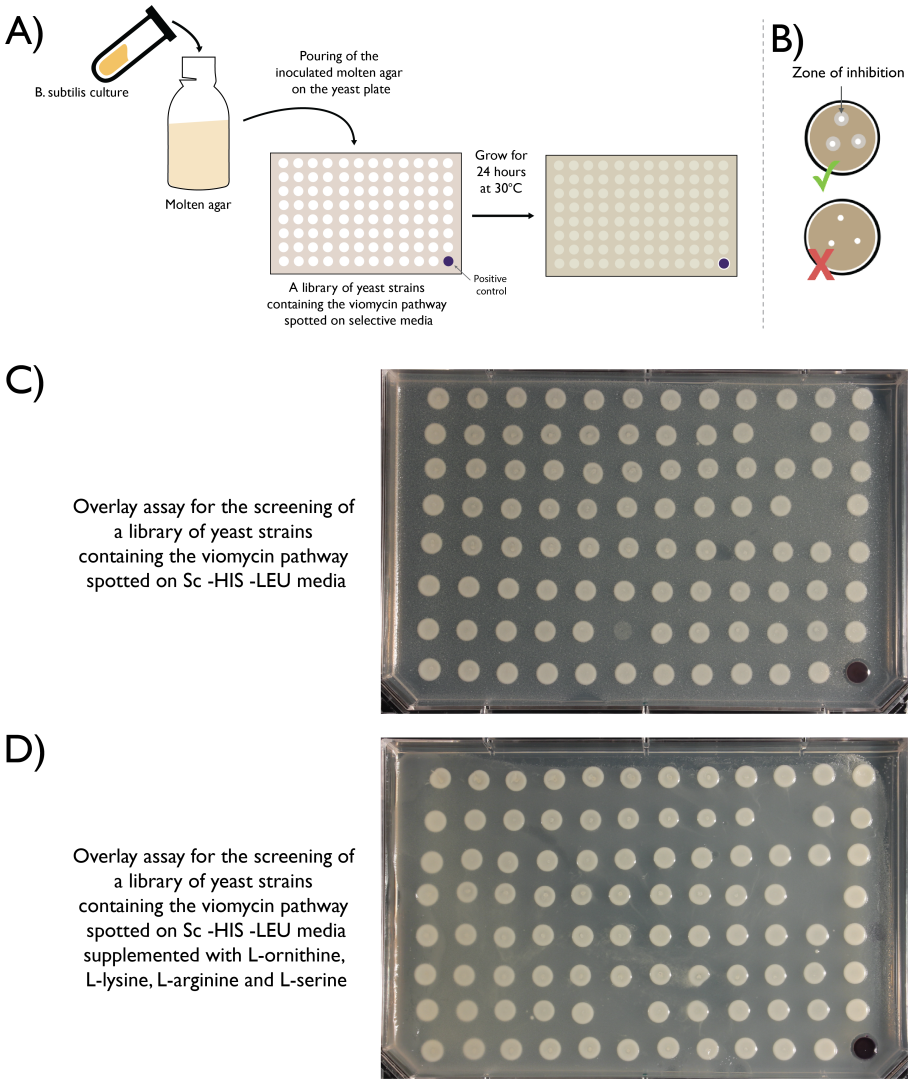


Figure 5.12

A) A diagram illustrating the overlay assay used to probe for the production of a bioactive compound against a tester strain (*B. subtilis*). In this assay, *B. subtilis* is added to molten overlay agar (48-50°C) then added on top of a yeast agar plate previously spotted with 95 strains containing the viomycin pathway (The dark strain producing violacein served as a positive control). The overlay agar is allowed to set before incubating it for 24 hours at 30°C. B) Yeast strains producing a antibiotic product will show a zone of inhibition or 'halo' surrounding the yeast colony C) The results for the *B. subtilis* overlay assay of 95 viomycin clones on SC -His -Leu media. The positive control for this assay is the growth of *B. subtilis* in violacein producing yeast strains. D) The results for the *B. subtilis* overlay assay of 95 viomycin clones on SC -His -Leu media supplemented with L-ornithine, L-lysine, L-arginine and L-seriine. The positive control for this assay is the growth of *B. subtilis* in violacein producing yeast strains.

5.2.5 Minimal viomycin core assembly

Difficulties in refactoring the viomycin pathway led to the truncation and the assembly of the minimal viomycin core. In this case, the enzymes involved in the tailoring of the core structure were removed, leaving only the 11 BSG genes involved in the construction of the viomycin cyclic pentapeptide core (Figure 5.13A). With the exception of *ppt*, all other viomycin genes were assembled with a promoter pool of moderate strength (Table 5.5). The *ppt* gene was put under the control of a strong promoter as its product is an important element in the functioning of NRPS's. The promoters, viomycin BSG genes and terminators were assembled into POT vectors according to the design plan described in Table 5.6 and subsequently transformed into *E. coli*.

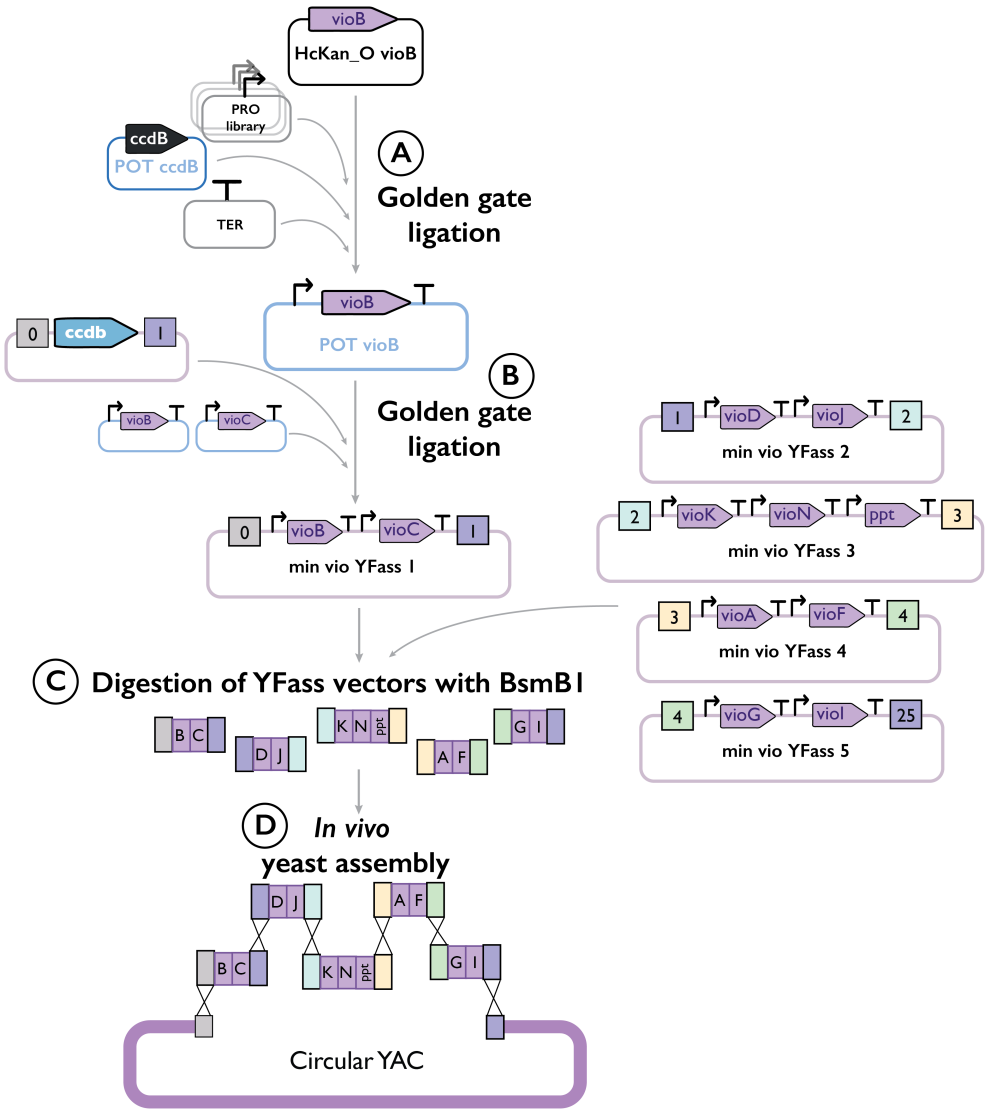


Figure 5.13

An overview of the minimal viomycin core assembly. A) Combinatorial assembly of a transcription unit, with a pool of promoters, a BSG gene and a terminator into a POT parts accepting vector by golden gate cloning. B) These transcription units can be assembled together (maximum of 6) into a multi-transcription unit vector (YFass) by another round of golden gate cloning. C) The assembled multi-transcription units are released from their respective YFass vectors by the digestion with BsmBI, revealing homology arms. D) These homology arms aid in the assembly of the full 11 gene pathway in a circular YAC by in vivo yeast homologous recombination.

Table 5.5 List of promoters used in the pool for the minimal viomycin core assembly

List of promoters in the moderate strength promoter pool	Size (bp)
pAAD6	500
pDIA4	288
pGPA1	500
pYPT1	350
pSBP1	500
pBCD1	245
pLSC2	446
pGRE3	437
pSBE22	500
pFSH1	500
pRPL2A	500
pELP2	500
pVMA10	245
pPAU13	500
pRPN1	267
pUBP6	321
pSSZ1	391
pSLT2	455
pERG1	406
pNIC96	292
pPRP8	222
pERG11	500
pGCN20	500
pBAT1	500
pPUT2	217
pDED81	438

Table 5.6 Table of the promoters, terminators and POT vectors used for the minimal viomycin core assembly. Each BSG was assembled with a pool of 3 promoters in a combinatorial assembly. The total size of each viomycin BSG assembly for all 11 genes was calculated (bp).

POT name	Promoter	ORF	Terminator	POT vector	Total insert size (bp)
min *vioB*_POT2	Promoter pool of moderate strength	*vioB*	tACS2	POT2	1704
min *vioC*_POT2		*vioC*	tENO2	POT3	1749
min *vioD*_POT2		*vioD*	tACS1	POT2	1925
minvioJ_POT2		*vioJ*	tCIT1	POT3	1928
minvioK_POT2		*vioK*	tFUM1	POT2	1838
minvioN_POT2		*vioN*	tACS1	POT4	855
minppt_POT2	pTDH3	*ppt*ase	tADH1	POT5	1427
min *vioA*_POT2	Promoter pool of moderate strength	*vioA*	tENO2	POT2	7238
minvioF_POT2		*vioF*	tCIT1	POT3	3885
minvioG_POT2		*vioG*	tACS2	POT2	4019
minvioI_POT2		*vioI*	tFUM1	POT3	2408

For each assembly reaction, all clones on the transformation plate were pooled instead of picking single colonies in order isolate maximal diversity from the POT assembly reaction. Due to high assembly success rates from previous assemblies, the POT assembly was proceeded to the next step of assembly. The appropriate POT vectors were assembled into five multi-transcription unit vectors (YFass 1-5 vectors) according to the design outlined in Table 5.7) and transformed into *E. coli* cells. Similarly to the previous viomycin POT assembly, the products of the combinatorial assembly were isolated by pooling all the colonies from the transformation plate. The combinatorial vio YFass vectors were digested with BsmBI and transformed along with the previously digested pYAC10 into yeast for the subsequent assembly of the viomycin pathway into a linear YAC by homologous recombination. As the left arm of the YAC possessed a HIS auxotrophic marker and the right YAC arm possessed a LEU marker, the yeast transformation was selected on SC -HIS -LEU -URA agar plates. Two minvio YAC clones were picked and screened for the presence of the full pathway by colony PCR (Figure 5.14A). Among them, minvio N6 (JAy021) appeared to possess the full pathway (Figure 5.14B).

Table 5.7 Assembly plan for the multi-transcription unit assembly for the minimal viomycin pathway and the calculated size of the final multi-transcription unit insert (bp)

YFass name	POT	YFass vector	Size of multi-TU insert (bp)
minvio_YFass 1	min *vioB*_POT2	YFass 1	3453
	min *vioC*_POT2		
minvio_YFass 2	min *vioD*_POT2	YFass 2	3853
	minvioJ_POT2		
minvio_YFass 3	minvioK_POT2	YFass 3	4120
	minvioN_POT2		
	minppt_POT2		
minvio_YFass 4	min *vioA*_POT2	YFass 4	11123
	minvioF_POT2		
minvio_YFass 5	minvioG_POT2	YFass 5	6427
	minvioI_POT2		

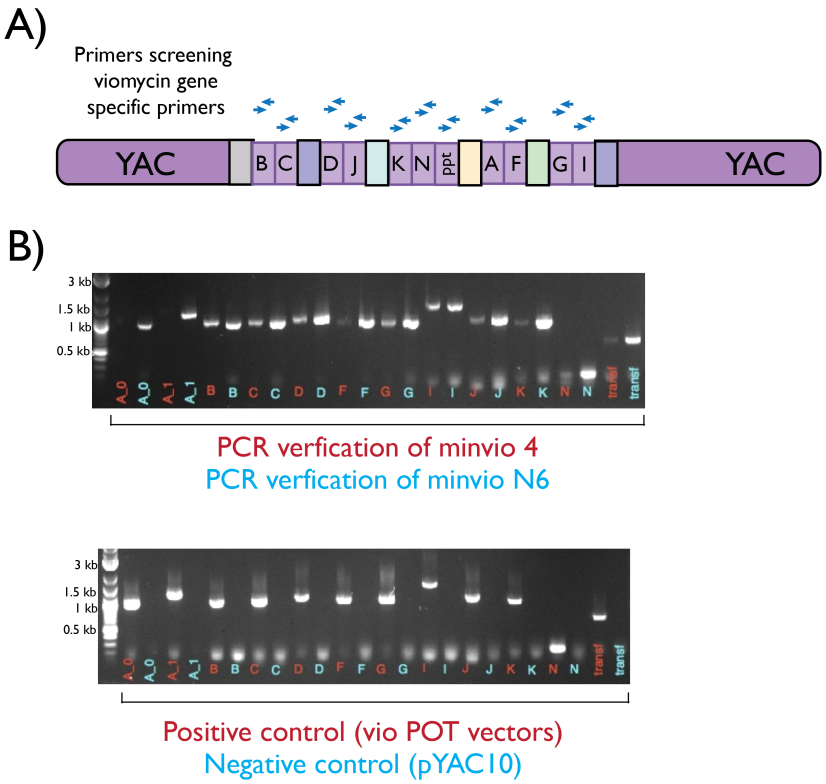


Figure 5.14

A) Diagrammatic representation of the primers used for the full pathway screening of 2 clones containing the minimal viomycin core B) Gel images for the colony PCR verification of the 2 clones from the minimal viomycin core assembly, alongside positive and negative controls.

Cell-based screening of JAy021 was subsequently carried out, using the overlay assay described earlier. The violacein strain (JAy019) sered as a positive control for the assay and the strain containing the empty YAC vector (JAy096) served as a negative control. JAy019 with *B. subtilis* showed a zone of inhibition or 'halo' surrounding the yeast colonies on the plate (Figure 5.15A). This 'halo' was not observed in the negative control JAy096 strain.

JAy021 was also screened against *Mycobacterium smegmatis* using the overlay assay, however due to the agglutination of *M. smegmatis* when grown in liquid culture, it was difficult to evenly distribute the microorganism across the entire overlay agar. Therefore, the protocol was modified as described in Materials and Methods section. After modification of the protocol, a faint 'halo' was observed in JAy021, which was not present in the negative control plate containing JAy096(Figure 5.15B).

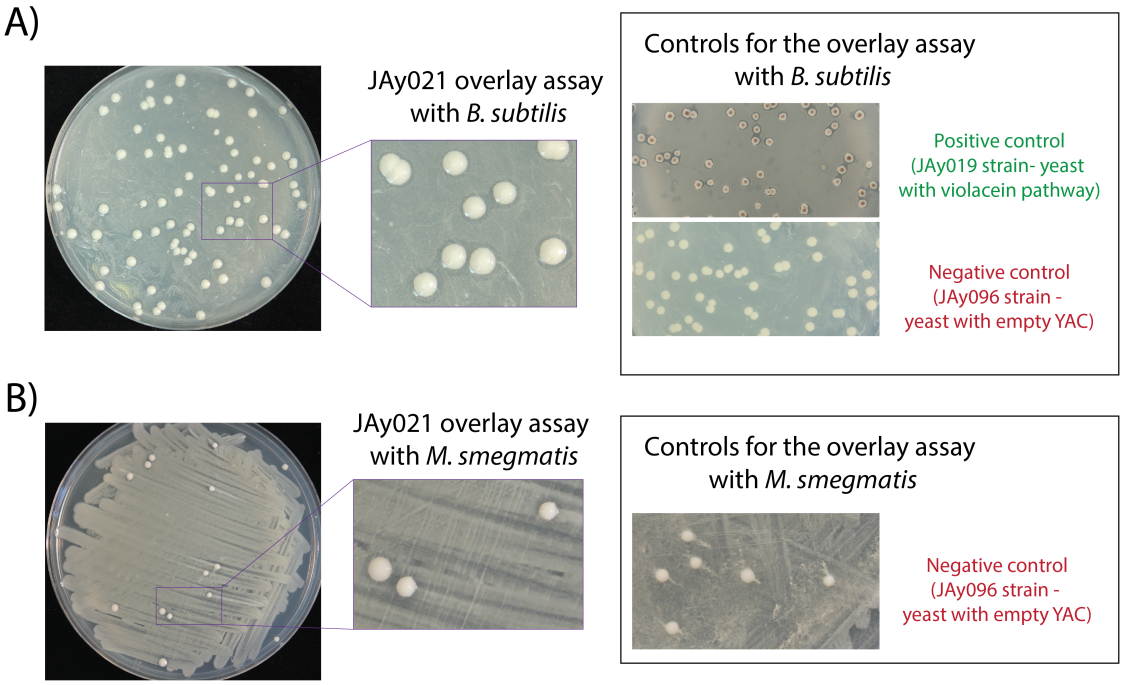


Figure 5.15

A) The results for the JAy021 overlay assay screening against *B. subtilis*, showing a zone of inhibition surrounding the yeast colonies. The positive and negative controls for this assay are violacein producer and a yeast strain containing an empty YAC respectively. B) The results for the JAy021 overlay assay screening against *M. smegmatis*, showing a zone of inhibition surrounding the yeast colonies. The negative control for this assay is a yeast strain containing an empty YAC. The negative control for the assay does not show a zone of inhibition surrounding the yeast colony

Dr. Venkateswarlu Yarlagadda, Dr. Grace Yim and Elizabeth Culp from the Wright lab at McMaster University further analysed the JAy021 strain to verify the findings observed in Figure 5.15B. They lysed and extracted compounds produced in the JAy021 producer strain and the negative control strain (JAy096) according to the protocol described in Materials and Methods - Viomycin extraction and screening by LC-MS. They observed antibiotic activity against *M. smegmatis* in the cell pellet fraction of JAy021. No bacterial inhibition was observed in the JAy021 spent media or the cell pellet fraction of JAy096 negative control (Appendix figure E.8A). Comparative analysis of the JAy096 negative control cell pellet fraction with cell pellet fraction of JAy021 by LC-MS was carried out by Dr. Yarlagadda. He identified a distinct peak with a mass to charge ratio (m/z) of 253.32 that was only present in the JAy021 cell pellet (Appendix figure E.8B). Further analysis by nuclear magnetic resonance (NMR),

performed by Dr. Yarlagadda, revealed that the identity of the anti-mycobacterial compound produced was palmitoleic acid (data not shown). This compound is a yeast metabolite and a fatty acid with known antimycobacterial activity.

5.2.6 Analysis of biosynthetic enzyme production by Western blot

All previous attempts for the refactoring of the viomycin pathway in yeast have been unsuccessful. Therefore, the first step in troubleshooting was the verification of protein production of all 20 genes by western blot. Viomycin BSGs were assembled with the pZ3EV promoter and tACS2 terminator into the POT1-*ccd*B vector by golden gate assembly and transformed into *E. coli*. One clone was verified by PCR for each of the 20 genes and each were subsequently transformed into the SASy31 strain. All 20 yeast strains were grown in 5 ml of SC -URA media to an OD600 of 0.5. The total protein of each strain was extracted from 2 ml of each culture at exponential phase using the protocol described in Materials and Methods section. The remaining culture was left to grow up till saturation (to an OD600 of ∼8) where another 2 ml was used for total protein extraction. All whole cell lysates were run on gradient PAGE gels (2-15%) and transferred onto nitrocellulose membranes. As every viomycin enzyme contains a FLAG tag on the C-terminal end, anti-FLAG antibodies can be used to detect the presence of each viomycin biosynthetic enzyme produced in each strain. FLAG tagged proteins were visualized by chemiluminescence and imaged using a Chemidoc MP. To account for differences in the amount of protein loaded, the total protein on the gel was imaged using the Biorad stain free technology using the Chemidoc MP (Biorad) (Appendix Figure E.9). The bands on the western blot were compared to their predicted sizes as outlined in Table 5.8. All viomycin enzymes appear to be produced in the western blot except for VioA, VioE, VioN, and VioP (Figure 5.16). A common feature across all cell lysates extracted at exponential phase are bands present due to binding of the anti-FLAG antibody to a protein present in the yeast at exponential phase but not in stationary phase (Figure 5.16A and 5.16B). The sizes of these bands are 250, 130, 90 and 40 kDa.

Despite all viomycin genes being driven by the same promoter, low expression was observed with Vph and VioA. However, it is difficult to distinguish VioA and the band present in all cell lysates taken at exponential phase (with a size of ∼250 kDa) due to

their size similarities (Figure 5.16A). This VioA band is not present in stationary phase, where the band due to non-specific binding of the antibody is also missing, therefore making it difficult to determine whether there is a problem in VioA expression (Figure 5.16B). Future work can focus on optimising the Western Blot protocol, verifying using another tag such as myc or His or including multiple copies of the FLAG tag to amplify the signal.

No production for VioE, VioN and VioP was observed in both conditions (exponential phase and stationary phase) on the Western blot (Figure 5.16A and 5.16B). Although VioE and VioP were not necessary in the production of the minimal viomycin core, VioN is an essential enzyme as well as Vph and VioA. No band may have been observed with VioN due to its small size ($\sim$8 kDa).

A method to investigate this is to use a higher percentage acrylamide gel to obtain better visualisation of smaller proteins. Alternatively, VioN can be fused with a fluorescent protein to allow the analysis of the strain by multiple methods (microscopy, fluorescence measurement on a plate reader or using a western blot). Another possibility is that the C-terminal FLAG tag was inaccessible due to specific protein folding. Therefore, further construction and screening of VioA, VioE, VioN and VioP can be used to confirm results observed in this study.

Table 5.8 Sizes of protein in kDa for each viomycin biosynthetic enzyme.

ORF name	Protein size (kda)
vioA-FLAG	229.63
vioB-FLAG	38.49
vioC-FLAG	40.43
vioD-FLAG	43.05
vioE-FLAG	47.67
vioF-FLAG	116.32
vioG-FLAG	116.91
vioH-FLAG	29.75
vioI-FLAG	60.88
vioJ-FLAG	42.83
vioK-FLAG	38.89
vioL-FLAG	33.71
vioM-FLAG	51.45
vioN-FLAG	7.98
vioO-FLAG	65.29
vioP-FLAG	51.41
vioQ-FLAG	39.12
vioS-FLAG	48.32
vph-FLAG	32.11
pptase-FLAG	27.13

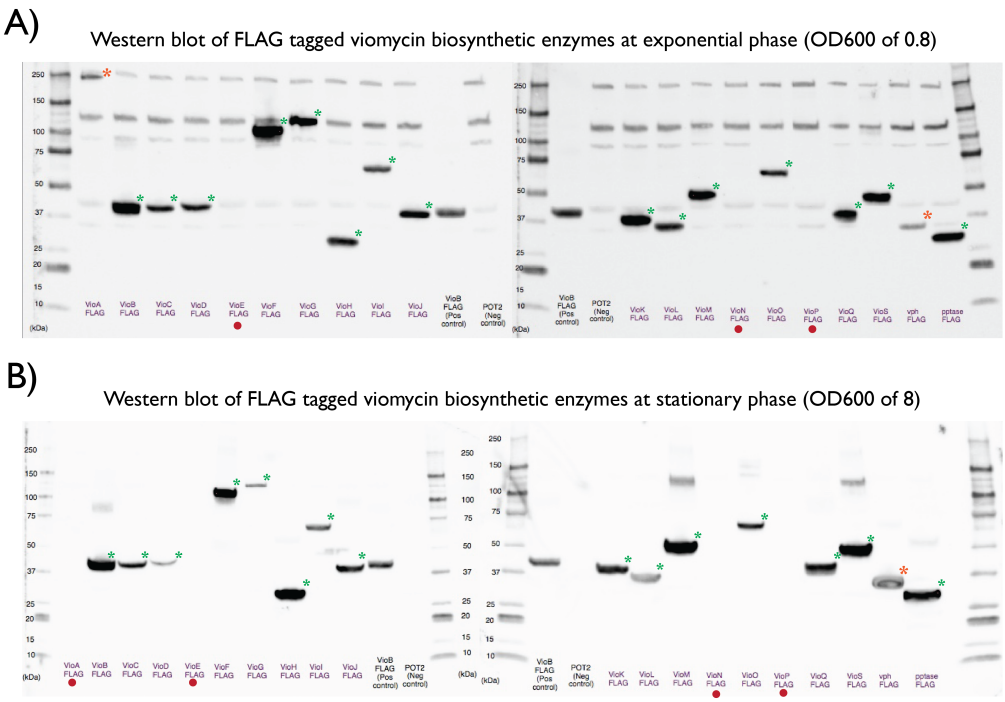


Figure 5.16

Western blot of all 20 viomycin biosynthetic enzymes in A) exponential phase and B) in stationary phase.

5.3 Chapter 5 discussion

This body of work demonstrates the capability of the expanded YeastFab cloning pipeline for the assembly of a large 20 gene viomycin pathway in yeast. Screening of the clones obtained at the end of the cloning process showed ∼80% assembly efficiency of the full pathway. However, despite evidence showing the expression of all 20 genes in one clone (JAy001) by RT-PCR, the presence of viomycin or its intermediates were not within detectable limits.

The effect of modulating viomycin gene expression on compound production was tested by combinatorially assembling each BSG with a pool of promoters with different activities. Screening at the end of the cloning process contained all 20 genes. The lower efficiency seen with combinatorial assembly relative to the single promoter assembly can be attributed to misassemblies as the digestion verifications experiments within the

cloning process showed bands not conforming to the predicted sizes, possibly due to the reassembly of the *ccd*B cassette into the acceptor vector. Another possible cause of the mis-assembly is the presence of BsaI and BsmBI sites remaining in the backbone vector. All these misassemblies still contain internal BsaI and BsmBI sites, which the final assembly does not contain. Therefore, a method to increase the assembly efficiency is by introducing a digestion step at the end of the cycling golden gate program to linearise all plasmids containing internal BsaI or BsmBI sites followed by the specific removal of linear double-stranded DNA using PlasmidSafe ATP dependent DNase (Epicentre).

Troubleshooting of the viomycin pathway through the expression of the minimal set of enzymes for the production of the cyclic pentapeptide resulted in a strain producing an antimycobacterial compound. Further analysis of the identity of this antibiotic compound showed it to be palmitoleic acid. As this is also a fatty acid yeast metabolite, it is possible that expression of the viomycin BSG resulted in broader metabolic changes. One hypothesis that the overexpression of 4'-phosphopantetheinyl transferase (PPTase), by putting it under the control of strong pTDH3 promoter, resulted in the high production of palmitoleic acid. Phosphopantetheinyl transferases are a group of enzymes involved in the production of fatty acids, nonribosomal peptides and polyketides (Copp and Neilan, 2006). Although the Pptase used in this study is from *B. subtilis*, it may be functional in yeast with sensitivity to endogenous metabolites that it can subsequently convert resulting in an increase in fatty acids. However, to confirm this, further analysis on the transcriptome and proteome will need to be carried out. Another method to confirm this hypothesis is to clone the PPTase with an inducible promoter (ie. pZ3EV or pGAL1) into a POT vector and transform it into yeast. This strain can subsequently be screened for an increase in palmitoleic acid.

Difficulties in producing viomycin or its intermediates led to the analysis of the individual enzymes. In order to achieve this, each enzyme was individually assembled with the same promoter (pZ3EV) and terminator (tACS2) combination into the POT1-*ccd*B vector. Each viomycin POT was separately screened for their expression by western blot and it was revealed that several enzymes were not expressed (VioE, VioN and VioP), with one key NRPS enzyme, VioA, appearing to only be expressed in a low level at exponential phase. However, further verification by alternative methods need to be carried out to validate this result. Another option is the use of alternative enzyme variants (from different native hosts) however due to limited funding for DNA synthesis, this was not tested. Future work can utilise the expanded YeastFab assembly

method for the combinatorial assembly of different enzyme variants (from different native hosts or using different codon optimisation parameters etc.).

Traditional metabolic engineering efforts rely on incremental one-factor-at-a-time (OFAT) method and building up the pathway, one enzyme at a time. However, this method can be very time consuming and expensive to troubleshoot, especially with large pathways like viomycin. Therefore, in this method, we relied on the large-scale construction of yeast strains containing different variants of the viomycin pathway, with different promoters driving each of the genes. The expanded YeastFab method also allows the combinatorial assembly of different parts allowing the screening of multiple factors such as promoters, terminators, enzyme variants and order of each gene on the plasmid. However, the work described above only included the screening of promoter choice for effects on viomycin expression. This library was then analysed using high-throughput screening of strains for the production of an antimicrobial compound. The screening method utilised in this pipeline is an overlay assay where the outcome of the results can be affected by multiple factors such as the amount of bioactive compound produced, thickness of the overlay agar and whether the antibiotic compound was exported. If any antibiotic compound being produced in the yeast cell did not reach the minimum inhibitory concentration (MIC), this could result in a false negative result. Similarly, if the compound was being produced but was not exported out of the cell, this strain would be scored as a false negative. A possible optimisation of the screening process could be achieved through the incorporation of a viomycin-sensitive riboswitch within the producer strain, allowing another method of screening to provide more robust results. Additionally, *E. coli* biosensors sensitive to inactive viomycin intermediates can be used to screen the spent media of yeast strains containing the pathway to help identify key bottlenecks.

Chapter 6

Construction and high-throughput screening of naringenin producer strains

6.1 Chapter 6 introduction

Flavonoids are a large family of plant metabolic intermediates that have attracted a lot of interest due to their beneficial properties as antioxidants, anti-inflammatory, antiviral and anti-tumour drugs (Nijveldt et al., 2001). Flavonoids are derived from activated phenylpropanoid precursors along with three molecules of malonyl-CoA. Examples of phenylpropanoids are phenolic acids such as *p*-coumaric acid, ferulic acid and caffeic acid, all of which function as precursors for the production of plant natural products such as lignins, coumarins and flavonoids (Weisshaar and Jenkins, 1998).

Naringenin is an essential scaffold flavonoid that can be further modified to produce a multitude of other compounds from different flavonoid classes such as flavones, isoflavones, flavonols, catechins and anthocyanins. Naringenin production begins with the deamination of L-phenylalanine mediated by phenylalanine ammonia lyase (*PAL*) resulting in the production of trans-cinnamic acid. Cinnamate-4-hydroxylase (*C4H*), a cytochrome P450 dependent enzyme, hydroxylates trans-cinnamic acid at the para position of the benzyl ring resulting in the formation of *p*-coumaric acid. Subsequently, *p*-coumaroyl:CoA ligase (*4CL*) activates *p*-coumaric acid producing *p*-coumaroyl-CoA.

Chalcone synthase (*CHS*), a type II polyketide synthase, then catalyses the sequential addition of three acetate extender units to an activated *p*-coumaroyl-CoA starter unit, utilising three molecules of malonyl-CoA. This produces naringenin chalcone or tetrahydroxychalcone. Finally, naringenin, a three ringed flavonone structure, is formed from naringenin chalcone through the action of chalcone isomerase (*CHI*) (Figure 6.1) (Watts et al., 2004). This conversion has also been shown to occur spontaneously (Mol et al., 1985).

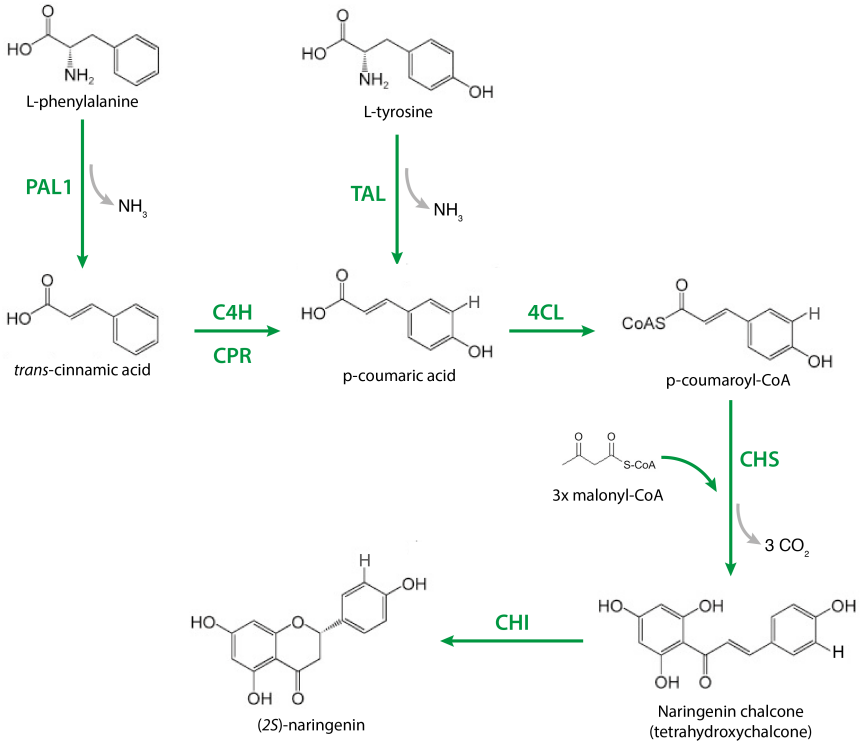


Figure 6.1

Naringenin biosynthetic pathway. In green are the enzymes involved in catalysing the reaction.(Koopman et al., 2012)

Currently, the main method of isolating flavonoids such as naringenin is through extraction from plants. However, the production of these compounds in plants is limited due to low natural yields. In addition, the slow growth rate of the plants and the complexity of the extraction process further limit the yield. An alternative method is to chemically produce the compound, however chemical production is restricted due to its relatively complex chemical structure. As a result, researchers have utilized

metabolic engineering to introduce the pathway into microorganisms such as *E. coli* and yeast.

Several publications have reported the engineering of both *E. coli* and *S. cerevisiae* to convert phenylpropanoid precursors such as *p*-coumaric acid into naringenin. Xu *et al.* have successfully engineered *E. coli* to produce high titres (474 mg/L) of naringenin from 2.5 mM *p*-coumaric acid (Xu et al., 2014). Yan *et al.* have engineered yeast to produce 28 mg/L of naringenin from 1 mM *p*-coumaric acid (Yan et al., 2005). A disadvantage of these engineered strains is the cost of the *p*-coumaric acid fed to produce naringenin.

Addressing this limitation, Koopman *et al.* engineered *S. cerevisiae* to produce the key intermediate flavonoid, naringenin, solely from glucose and without the need for supplementation of the growth media phenylpropanoid precursor (Koopman et al., 2012). All biosynthetic enzyme genes (*atPAL*, *atC4H*, *at4CL*, *atCH3* and *atCHI1* – 'at' for *Arabidopsis thaliana*) used for the production of naringenin from phenylalanine were from *Arabidopsis thaliana*. Their decision to obtain biosynthetic enzyme genes from a single donor organism was due to observed protein-protein interactions. Deriving all enzymes from the same donor can lead to potentially improved efficacy of functionality due to improved kinetic relationship and enzyme specificity (Winkel-Shirley, 1999). Initial construction of a naringenin producer strain by Koopman and colleagues resulted in low extracellular naringenin yield from phenylalanine (1.5 mg/L) (Koopman et al., 2012). The introduction of the heterologous gene for tyrosine ammonia lyase (*TAL*) from *Rhodobacter capsulatus* allowed the deamination of tyrosine to provide an alternative route to *p*-coumaric acid and ultimately improved naringenin yields to 54.45 mg/L (Koopman et al., 2012). Similarly, Lyu *et al.* engineered *S. cerevisiae* for the production of naringenin, however they focussed on the utilisation of only tyrosine for naringenin production. Unlike Koopman et al, Lyu and colleages obtained pathway genes from various hosts (TAL from *Flavobacterium johnsoniae*, *4CL* from *A. thaliana*, *CHS* from *Hypericum androsaemum* and *CHI* from *Medicago sativa*) (Lyu et al., 2017).

6.1.1 Improving the flux of precursor molecules

Further optimization of the yeast titres involved modification of the yeast metabolism, altering the flux towards improving precursor supply. Naringenin utilises tyrosine and phenylalanine as base compounds for its synthesis. Biosynthesis of aromatic amino

acids, such as phenylalanine and tyrosine, in yeast is regulated by feedback inhibition of 3-deoxy-D-arabinose-heptulosonate-7-phosphate (DAHP) synthase (*ARO4* and *ARO3*) and chorismate mutase (*ARO7*) (Luttik et al., 2008). As such the negative feedback mechanism prevents high accumulation of the precursors within the cell.

Luttik *et al.* showed that the point mutation of *ARO4* ($ARO4^{K229L}$) and *ARO7* ($ARO7^{G141S}$) altered their enzymatic activity, disrupting the feedback inhibition and resulting in a 200-fold increase in aromatic compound accumulation (Luttik et al., 2008).

Koopman *et al.* employed a similar strategy to disrupt this feedback inhibition by deleting *ARO3* and introducing a tyrosine insensitive *ARO4* (Koopman et al., 2012). Elimination of this feedback inhibition resulted in the feedthrough of phenylalanine to the Ehrlich pathway (Luttik et al., 2008). While this improved phenylalanine production, it also resulted in the accumulation of phenylethanol causing a reduction in growth, and this effect was mitigated by deleting the genes encoding phenylpyruvate decarboxylase (Aro10, Pdc5 and Pdc6). Combining all of these modifications within a single strain resulted in a 40-fold increase in naringenin yield (54 mg/L) in a shake-flask culture (Koopman et al., 2012). These optimisations by Koopman *et al.* have focussed on the metabolic engineering of the aromatic amino acid biosynthesis pathway, specifically on the phenylalanine route for naringenin production. In contrast, Lyu *et al.* focussed on optimising tyrosine availability for naringenin production. This was achieved by disrupting the feedback inhibition of DAHP synthase by overexpressing a mutated *ARO4* ($ARO4^{K229L}$) resulting in the shifting the metabolic flux towards tyrosine. This strain produced a higher naringenin titre to 8.9 mg/L compared to the original version (5.2 mg/L). Similar to the findings of Koopman et al, Lyu *et al.* found that all the modification resulted in the accumulation of phenylethanol causing a slight decrease in growth. Deletion of *PDC5* reduced phenylethanol accumulation and the recovery of the growth phenotype was observed. The final engineered yeast strain produced 84 mg/L of naringenin.

Described above are advances made in the field of metabolic engineering of *E. coli* and yeast for optimal naringenin production. Major developments centre around: the identification and compensation of rate limiting steps, removal of bottlenecks in pathway flux, elimination of competing pathways and modifying the metabolic flux for increasing the availability of precursor metabolites.

6.1.2 Development of biosensors

Advances in genome engineering over the past decade have allowed researchers to explore the effect of semi-rational genome-wide modifications on compound production. CRISPR and recombinases have allowed researchers to introduce a wide range of modifications in the yeast genome, several of which may result in the alteration of the yeast metabolic pathway, improving flux of precursors and ultimately improving compound yields. However, screening these large libraries of genetically diverse strains poses a challenge due to the reliance on relatively low-throughput and costly analytical techniques such as mass spectrometry. As such one key bottleneck is the capacity to obtain a reliable method for sifting through libraries of strains for the best producer.

Transcription factor-based biosensors responsive to intracellular small molecules offer a potential solution to this challenge. Many transcription factors have been identified in *E. coli* that are responsive to small molecules. This wide availability of proteins responsive to small molecules allows us to tap into this potential for the design of biosensors to enable high-throughput screening. The use of biosensors for screening can be applied for multiple purposes. Most commonly, biosensors have been used as a method to isolate high producers from a library of strains, each with a different modification that may or may not impact the production of a compound. Pfleger and colleagues have demonstrated the use of a biosensor for high-throughput screening of strains with the goal of optimising the production of mevalonate (Pfleger et al., 2006). Biosensors have also been used to detect levels of key intermediates and their temporal depletion (Xu et al., 2014). Skjoet and colleagues have also demonstrated the transplantation of prokaryotic biosensors into a eukaryotic chassis (Skjoedt et al., 2016). Trabelsi and colleagues sought to optimise titres of pinocembrin with the use of a biosensor to rapidly screen different versions of the pathway in a high-throughput fashion (Trabelsi et al., 2018).

Analysis of naringenin production in yeast by Lyu et al revealed that the intracellular naringenin concentration only amounted to 6-10% of the total naringenin produced (Lyu et al., 2017; Watts et al., 2004). Therefore, a biosensor for the detection of naringenin in the spent media is preferable. FdeR is a transcription factor from *Herbaspirillum seropedicae*, belonging to the LysR family of transcriptional regulators. FdeR triggers the activation of transcription for the gene cluster, *fdeABCDEFGHIJ*. Within this gene cluster, *fdeA* produces a protein that is critical for naringenin degradation (Figure 6.2). Siedler and colleagues have isolated this transcriptional regulator

(FdeR) and placed the *fdeA* promoter upstream of a CFP fluorescent protein (Figure 6.2) (Siedler et al., 2014). The expression of CFP was detected upon induction with 100 μM naringenin or 100 μM naringenin chalcone. They also observed a linear correlation between the amount of naringenin and the output fluorescence in concentrations from 5 μM to 100 μM naringenin. Saturation was observed between 100 μM and 300 μM naringenin (Siedler et al., 2014). Marin *et al.* showed that FdeR is not responsive to phenylalanine or tyrosine (Marin et al., 2013). However FdeR appeared to be sensitive to naringenin chalcone, naringenin and a related flavonoid, pinocembrin (Siedler et al., 2017; Trabelsi et al., 2018).

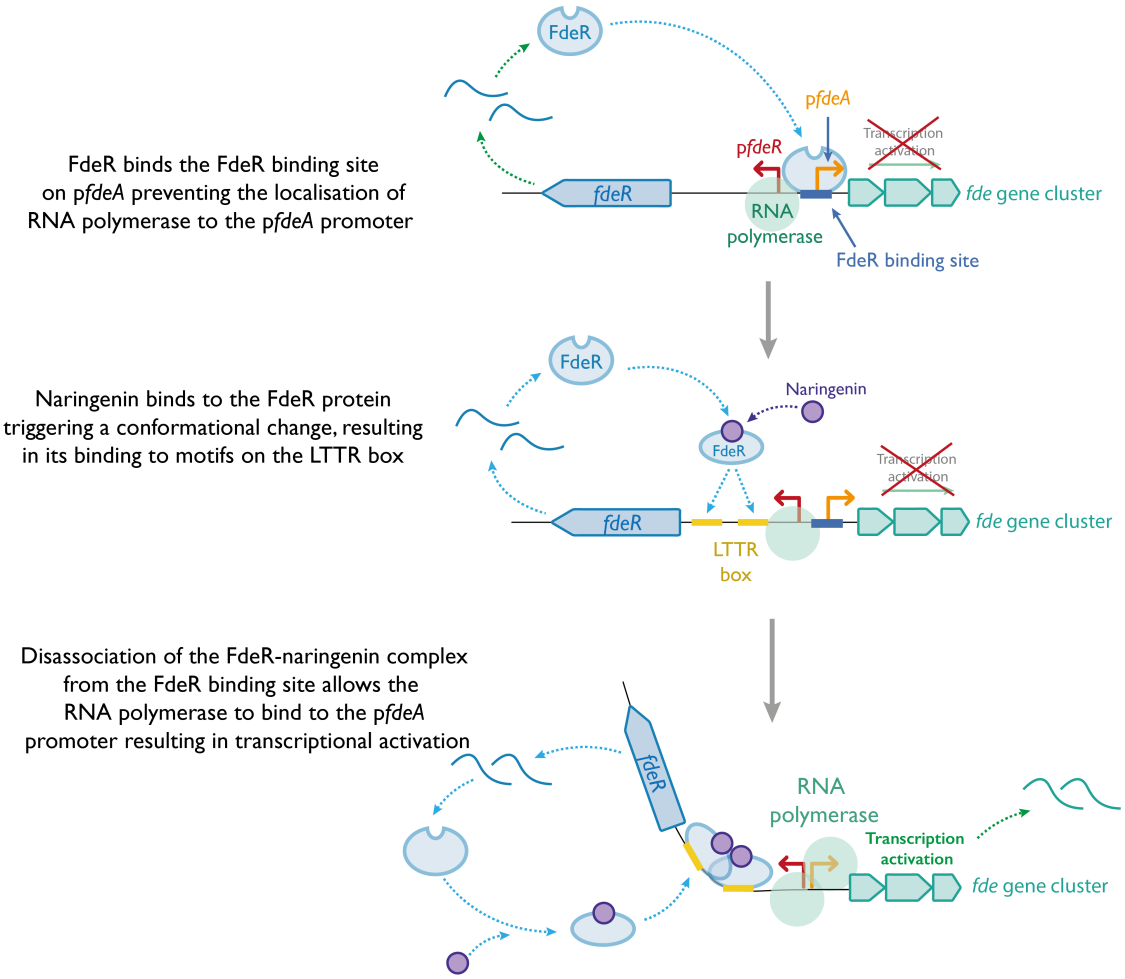


Figure 6.2

Mechanism of FdeR-mediated transcriptional regulation as described by Marin et al. (2013); Wassem et al. (2017)

This chapter begins with the assembly of the naringenin pathway using the YeastFab strategy. Vectors and regulatory elements (promoters and terminators) previously constructed as part of the toolkit were used for pathway assembly. The pathway is assembled in a hierarchical fashion beginning with the domestication of the naringenin biosynthetic enzyme genes (BSGs) into the YeastFab format (Figure 6.3). Subsequent optimisations on the pathway design are carried out to improve naringenin titres. This chapter then describes work carried out to develop a biosensor for naringenin in spent media to sift through libraries of genetically diverse strains for high producers. An *E. coli* biosensor is optimised for its use in a spent media assay for the detection of naringenin. This assay was used to determine the effects of 1) plasmid hosted versus genomically integrated pathways, 2) using different enzyme variants of *4CL*, and 3) using different yeast background strains on naringenin production.

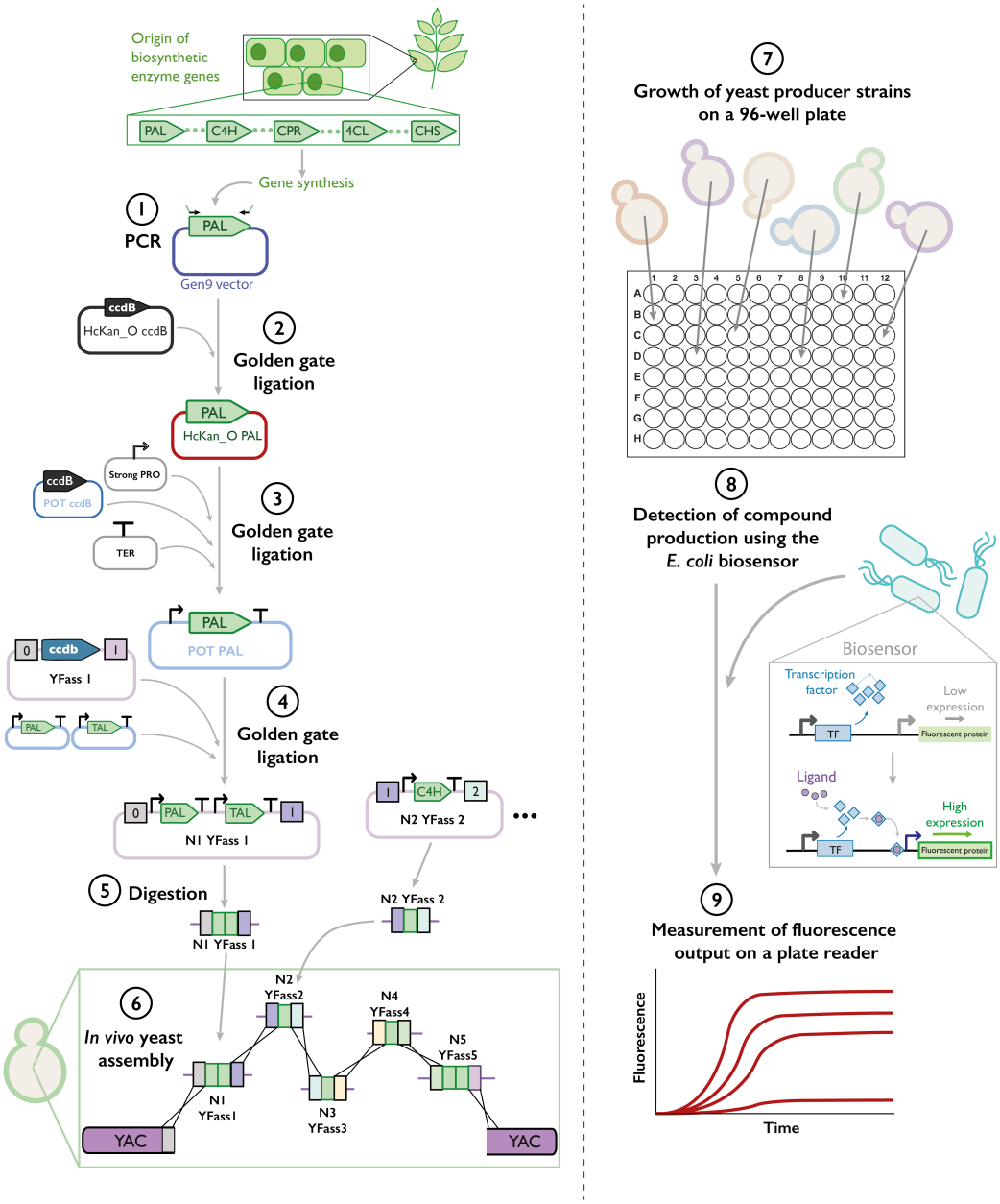


Figure 6.3

A) An overview of the naringenin pathway assembly using the YeastFab system. B) An overview of the spent media assay using an *E. coli*-based biosensor.

6.2 Chapter 6 results

6.2.1 Domestication of the naringenin biosynthetic enzyme genes to the YeastFab format

The naringenin biosynthetic enzyme genes (BSGs) were previously synthesised by a commercial vendor, Gen9 in a collaborative project (SynNP) with Tyers laboratory (University of Montreal) and Wright laboratory (McMaster University) (Appendix Table F. Initially, all BSGs were codon optimized for yeast expression and the BsaI, BsmBI and SapI restriction sites were recoded by Dr. Almer van der Sloot from the Tyers lab. However, these synthesised BSGs did not include BsaI type IIs restriction sites and appropriate overhangs for further pathway assembly using the YeastFab assembly method.

Therefore, I have designed primers to incorporate these elements (BsaI type IIs restriction sites and appropriate overhangs) flanking each naringenin BSG. On the forward primer, a FLAG tag was inserted immediately downstream of the start codon. This was inserted to allow downstream analysis of expression by Western blot. I amplfied the naringenin ORF parts and cloned them into the HcKan_O vector by Golden Gate assembly (Materials and Methods - PCR and Golden Gate assembly). One clone per BSG was further screened by colony PCR and sequence verified. All BSGs were successfully cloned with no mutations introduced (Appendix 6 - Figure F.1 and F.2).

6.2.2 Transcription unit assembly of the naringenin biosynthetic enzyme genes (POT assembly)

Each biosynthetic gene was subsequently assembled with a promoter and a terminator in a specific POT vector according to Table 6.1. The choice of which POT to use for each gene was dependent on the order of the subsequent YFass assembly. For this pathway, promoters characterised in the previous chapter (Chapter 3 – YeastFab Make) were used. The pathway was fragmented into two segments. The first segment consists of genes involved in converting aromatic amino acids into phenylpropanoid precursors (ie. *p*-coumaric acid). This includes *PAL*, *TAL*, *C4H* and *CPR*. The second

segment consists of genes involved in the conversion of *p*-coumaric acid to naringenin. This includes the *4CL*, *CHS* and *CHI* biosynthetic enzyme genes. Specific constitutive promoters were chosen and added to genes in the first segment (Table 1). The first two genes in the second segment, *4CL* and *CHS* were placed under the control of an inducible promoter (pZ_3EV). The final gene, *CHI*, was put under the control of pTDH3. In this round of assembly, the enzyme choice was based on those used by Lyu et al, the highest producer to date (Lyu *et al.*, 2017). With several enzyme variants for *CHI* in the SynNP library, multiple versions of the pathway were assembled to identify the optimal chalcone isomerase that can result in the highest amount of naringenin produced. Two colonies per assembly were screened by colony PCR using primers flanking the cloning site on the POT vector (POT chc FP and POT chc RP) (Table of primers). The calculated size of the amplicon (Table 6.1) and gel simulation was compared to the size of the band observed on the agarose gel following PCR verification. All assemblies possessed at least 1 clone with a similar band pattern to that found in the gel simulation and the correct length of amplicon (Figure 6.4A and B). One clone per assembly was plasmid purified for subsequent assembly into the multi-transcription unit vector.

Table 6.1
Assembly plan for POTs in the first round of naringenin pathway assembly

Name	Promoter	Promoter size (bp)	ORF	Gene length (bp)	Terminator	Terminator size (bp)	POT	Length of POT insert (bp)	Predicted PCR amplicon length (bp)
N1 POT	pRPS2	501	*atPAL1*	2202	tACS1	330	POT2	3033	3165
N2 POT	pACT1	501	*rcTAL*	1620	tCIT1	302	POT3	2423	2555
N3 POT	pRPS3	500	*atC4H*	1542	tRPS13	200	POT1	2242	2374
N4 POT	pPGK1	514	*scCPR1*	2100	tACS2	293	POT1	2907	3039
N5 POT	pZ$_3$EV	699	*at1-4CL3*	1512	tFUM1	288	POT1	2499	2631
N6 POT	pZ$_3$EV	699	*haCHS*	1197	tRPS3	200	POT2	2096	2228
C1 POT	pTDH3	518	*csCHI-1*	693	tENO2	294	POT3	1505	1637
C2 POT	pTDH3	518	*plCHI-2*	699	tENO2	294	POT3	1511	1643
C3 POT	pTDH3	518	*osCHI-3*	726	tENO2	294	POT3	1538	1670
C4 POT	pTDH3	518	*psCHI-4*	696	tENO2	294	POT3	1508	1640
C6 POT	pTDH3	518	*ph2CHI-6*	687	tENO2	294	POT3	1499	1631

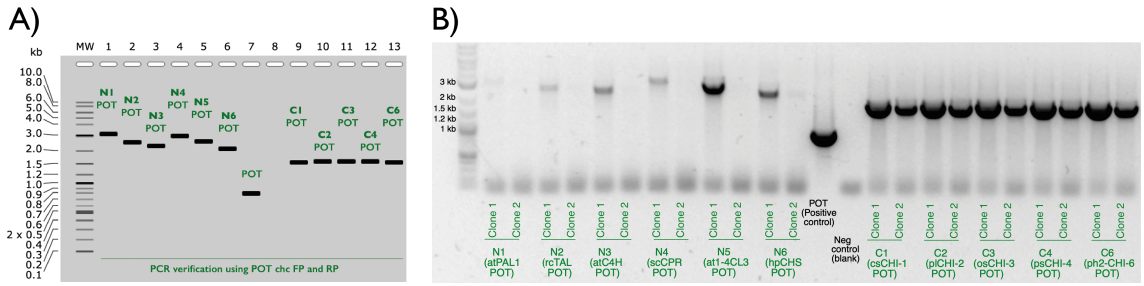


Figure 6.4

A) Gel simulation of PCR amplicons for POT verification. B) Agarose gel verification of PCR products

6.2.3 Protein expression validation by Western blot

POT vectors were then transformed into the SASy31 yeast strain (Table of yeast strains]) One clone per assembly was grown to an OD of 0.8 then subsequently lysed and analysed by Western blot, with the exception of the strain containing a naringenin biosynthetic enzyme gene driven by a β-estradiol inducible promoter (pZ_3EV), which was cultured in duplicate. One of the duplicates was induced with 1 μM of β-estradiol and the other was left uninduced. The cultures were lysed and a Western blot was performed, using anti-FLAG mouse primary antibody and anti-mouse HRP secondary antibody. FLAG tagged proteins were visualized by chemiluminescence and imaged using a Chemidoc MP (Materials and Methods section).

All enzymes appear to be expressed (Figure 6.5). Interestingly, the amount of protein detected on the western blot appears to be different between the *CHI* variants (*csCHI-1, plCHI-2, osCHI-3, psCHI-4* and *ph2CHI-6*) despite all the genes being driven by the same promoter (pTDH3) and terminator (tENO2). This could be due to several factors such as codon usage, RNA secondary structure, RNA stability, protein stability or protein toxicity. The highest expressed enzyme variant among these was *ph2CHI-6*. Similarly, *4CL* and *CHS* are driven by the same promoter, pZ_3EV, however the expression of *4CL* appears significantly less than that of *haCHS*. The expected sizes of the protein on the SDS-PAGE gel appear similar to their corresponding predicted sizes (Table 6.2) (Figure 6.5).

However, several other bands (three bands) were present in all lanes. These could possibly be off target binding of the primary or secondary antibody to yeast proteins.

It is however not present in the positive and negative control possibly due to protein degradation as these controls were previously extracted and used across all western blots in this study. This is further confirmed by the slightly fainter band pattern in the total protein image of the gel in Appendix figure F.3. The total protein on the gel was imaged using the Biorad stain free technology, showing similar amounts of protein lysate loaded for each strain (Appendix figure F.3). This suggests that differences between the *CHI* variants are not due to different amounts of total protein loaded on the gel.

Table 6.2

Predicted sizes of each naringenin BSG enzyme (in kDa) Labelled in green are N1-N6 POTs containing naringenin BSGs except for *CHI*. Each naringenin BSG is is assembled with a unique set of promoter and terminator. Labelled in yellow, orange, purple, violet and blue are 5 different variants of *CHI* tested, each assembled with the same set of promoter (pTDH3) and terminator (tENO2)

Strain name	POT plasmid construct information	kDa
N1 POT	pRPS2-*atPAL1*-tACS1 POT2	79.73
N2 POT	pACT1-*rcTAL*-tCIT1 POT3	56.54
N3 POT	pRPS3-*atC4H*-tRPS13 POT1	58.8
N4 POT	pPGK1-*scCPR1*-tACS2 POT1	77.75
C1 POT	pTDH3-*csCHI-1*-tENO2 POT3	24.98
C2 POT	TDH3_P-*plCHI-2*-tENO2 POT3	24.8
C3 POT	TDH3_P-*osCHI-3*-tENO2 POT3	24.89
C4 POT	TDH3_P-*psCHI-4*-tENO2 POT3	25.79
C6 POT	TDH3_P-*ph2CHI-6*-tENO2 POT3	24.94
N5 POT	pZ$_3$EV-*at1-4CL3*-tFUM1 POT1	53.3
N6 POT	pZ$_3$EV-*haCHS*-tRPS3 POT2	43.71

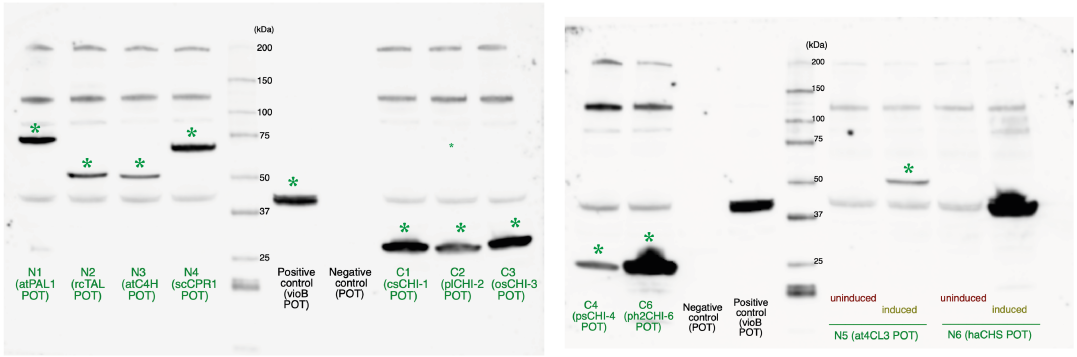


Figure 6.5

Western blot of total protein from strains containing naringenin BSGs assembled into a POT vector. Each gene was tagged with a FLAG tag and expressed from a unique strain. A green asterisk is used to label bands of interest on each lane.

6.2.4 Multi-transcription unit assembly of naringenin BSGs

Transcriptional units were subsequently cloned into the set of YFass 1-5 vectors (multi-transcription unit vectors) containing homology arms flanking the cloning site (Material and Methods - Golden Gate assembly). These homology arms allow the subsequent assembly of the full pathway into a circular plasmid or a linear YAC. For the assembly of 5 versions of the naringenin pathway, each having a different *CHI* enzyme variant, nine golden gate reactions were set up according to the table described below (Table 6.3).

Table 6.3

Assembly plan for multi-TUs (YFass) in the first round of naringenin pathway assembly. Labelled in yellow, orange, pink, purple and blue are 5 variants of the YFass5 assembly to test the difference in naringenin yields resulting from the different variants of *CHI*.

YFass assembly name	POT plasmid name	POT plasmid construct information	YFass vector
N1 YFass1	N1	pRPS2-*atPAL1*-tACS1 POT2	YFass 1
	N2	pACT1-*rcTAL*-tCIT1 POT3	
N2 YFass2	N3	pRPS3-*atC4H*-tRPS13 POT1	YFass 2
N3 YFass3	N4	pPGK1-*scCPR1*-tACS2 POT1	YFass 3
N4 YFass4	N5	pZ$_3$EV-*at1-4CL3*-tFUM1 POT1	YFass 4
N5-C1 YFass5	N6	pZ$_3$EV-*haCHS*-tRPS3 POT2	YFass 5
	C1	pTDH3-*csCHI-1*-tENO2 POT3	
N5-C2 YFass5	N6	pZ$_3$EV-*haCHS*-tRPS3 POT2	YFass 5
	C2	TDH3_P-*plCHI-2*-tENO2 POT3	
N5-C3 YFass5	N6	pZ$_3$EV-*haCHS*-tRPS3 POT2	YFass 5
	C3	TDH3_P-*osCHI-3*-tENO2 POT3	
N5-C4 YFass5	N6	pZ$_3$EV-*haCHS*-tRPS3 POT2	YFass 5
	C4	TDH3_P-*psCHI-4*-tENO2 POT3	
N5-C6 YFass5	N6	pZ$_3$EV-*haCHS*-tRPS3 POT2	YFass 5
	C6	TDH3_P-*ph2CHI-6*-tENO2 POT3	

Four clones were isolated from the transformation plate and individually screened for the presence of each naringenin BSG in the extracted plasmid by PCR verification using naringenin Golden Gate cloning primers (Table of primers) (Figure 6.6A). Amplicon sizes for each PCR product showed a similar pattern to that shown in the gel simulation, except for *PAL* and *CPR*, both of which show faint bands of the correct size along with other bands of smaller sizes (less than 500 bp) (Figure 6.6A and B). Additional verification by restriction digestion a pattern matching that of the predicted digestion simulation (Figure 6.6C and D).

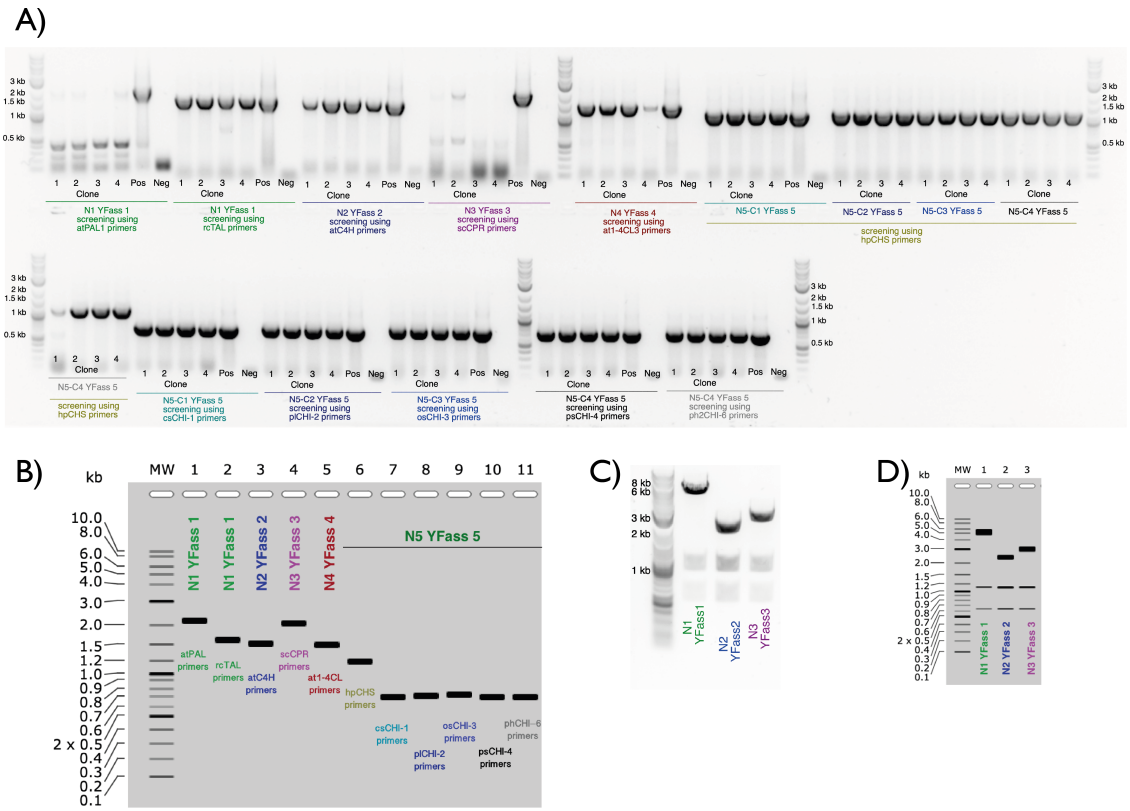


Figure 6.6

Verification of naringenin YFass assemblies. A) Agarose gel of the PCR products from the verification of YFass multi-TU assemblies. B) Simulation of amplicons from the PCR verification of YFass multi-TU assemblies. C) Digestion verification using BsmBI of YFass 1, YFass 2 and YFass 3. D) Digestion simulation of YFass 1, YFass 2 and YFass 3 to confirm results from the digestion verification

6.2.5 Full pathway assembly of the naringenin pathway

YFass vectors were purified and digested using BsmBI to release the transcription units or multi-transcription units along with the appropriate homology arms from the YFass vector. The pYAC10 plasmid was digested using XbaI and SacI to linearise the vector, revealing the homology arms to facilitate subsequent recombination. These digestion reactions were column purified and transformed into yeast (Materials and Methods - Yeast transformation). Transformants were grown in SC -HIS -URA dropout

medium to select for strains harbouring the circularized pYAC10 plasmid (Figure 6.7A).

One clone was isolated from the transformation plate and grown for storage in glycerol at -80℃ and verification experiments. The strain ID and a diagrammatic representation of each pathway in each strain constructed is illustrated in (Figure 6.7B). All four strains were grown and subsequently verified for the presence of all naringenin BSGs by PCR (Materials and Methods - Yeast genomic DNA extraction and PCR). PCR verification for each gene using the naringenin Golden Gate cloning primers was carried out (Table of primers). All genes appeared to be present in NC6, NC2, NC3 and NC4, with bands produced at the expected size and no bands obtained in the negative control (YAC10) (Figure 6.7C and D). An initial Sanger sequencing verification of 1 strain (JAy081) was carried out by primer walking. Appendix figure F.4 shows the correct assembly of all 7 genes into the pYAC10 vector and no introduction of mutations into any of the BSGs.

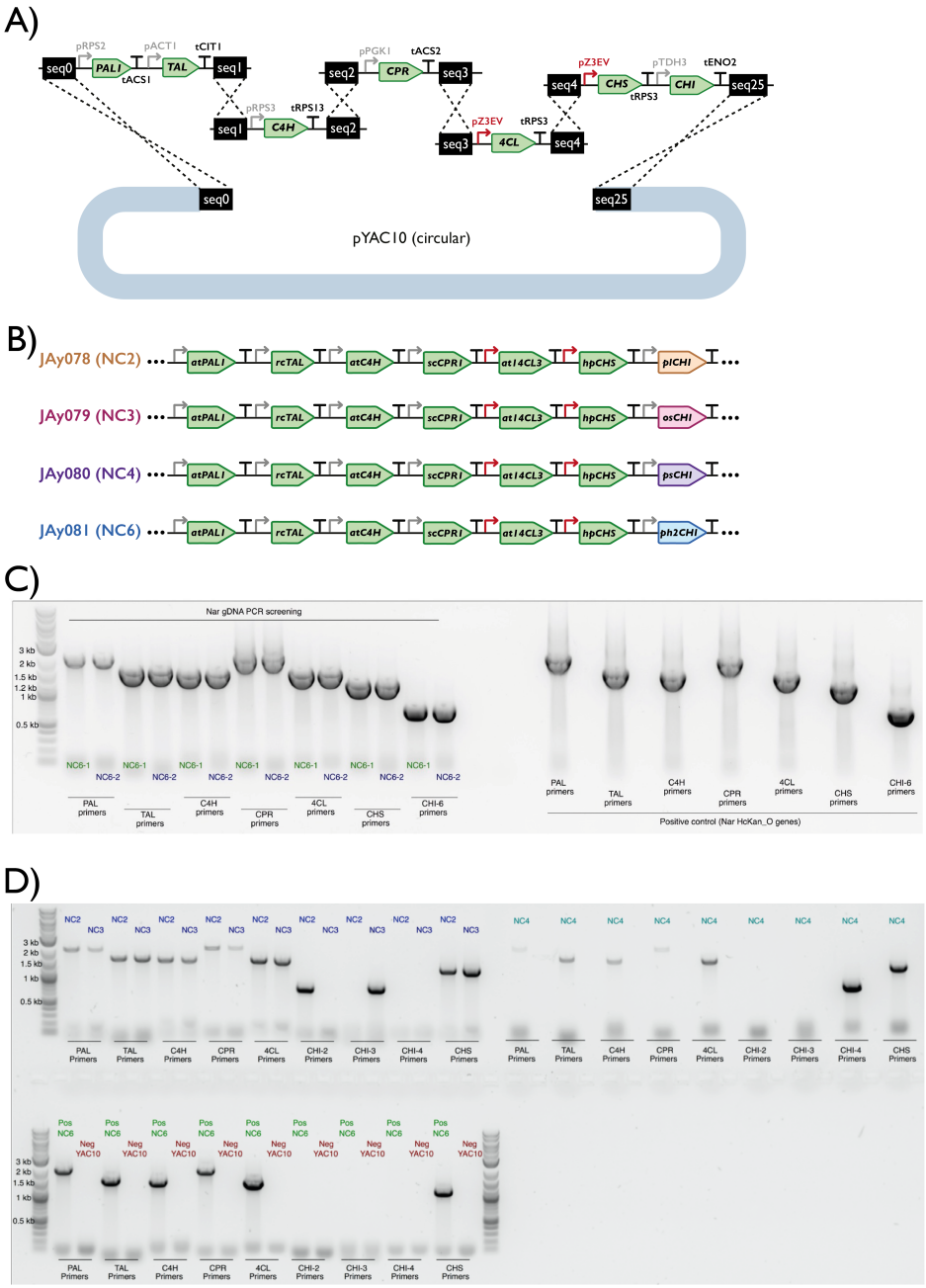


Figure 6.7

Assembly of the naringenin pathway into the YAC10 vector. A) Diagrammatic representation of the naringenin pathway assembly into the YAC10 vector by homologous recombination. B) Strain ID and a diagrammatic representation of the final naringenin pathway variants. The promoter labelled in red is an inducible promoter (pZ_3EV). Promoters in grey are constitutive promoters. C) PCR verification of JAy081 (NC6) for the presence of all 7 naringenin genes. D) PCR verification of JAy078 (NC2), JAy079 (NC3), JAy080 (NC4) for the presence of all 7 naringenin genes.

6.2.6 Characterization of *p*-coumaric acid and naringenin production in JAy078, JAy079, JAy080 and JAy081 strains

Initial analysis of two strains (JAy079) containing the naringenin pathway was carried out (Materials and methods - naringenin chemical analysis). Three biological replicates were grown overnight in SC-His-Ura medium with 1 μM β-estradiol. Following 24 hours of growth, the cell culture was screened for the amount of *p*-coumaric acid and naringenin in the spent medium using the mass spectrometer, with help on sample processing on the mass spectrometer and data analysis from Dr. Cunyu Yan. While no naringenin was detected, however, *p*-coumaric acid was observed (Figure 6.8).

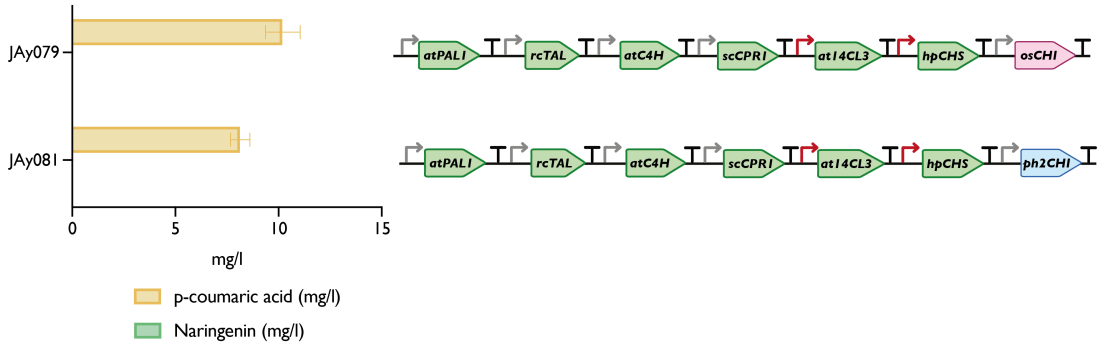


Figure 6.8

Quantification of *p*-coumaric acid and naringenin production using a mass spectrometer (n=3).

From these results, it appears that a low amount of *p*-coumaric acid is being produced, suggesting that the first segment of the pathway (comprising of *PAL*1, *TAL*, *C4H* and *CPR*) is functional. To test for the activity of later stages of the pathway, I utilised the same strategy utilised by Watts *et al.* for the troubleshooting of naringenin production (Watts et al., 2004). In their publication, a high amount of phenylpropanoid precursor (*p*-coumaric acid) was introduced to the medium prior to growth. The excess of *p*-coumaric acid can hopefully trigger production of naringenin or aid in the identification of non-functional enzymes (Figure 6.9A).

Two biological replicates of JAy078, JAy079, JAy080 and JAy081 were grown overnight in SC-His-Ura medium, spiked with 1 μM β-estradiol and 150 μM (26.606 mg/L) *p*-coumaric acid. After 24 hours of growth, the culture was prepared for analysis using the mass spectrometer for analysis of *p*-coumaric and naringenin production in technical triplicates, with help on sample processing on the mass spectrometer and

data analysis from Dr. Cunyu Yan. *p*-coumaric acid was detected in the spent media in similar titres to those with which the culture was spiked at the beginning of the experiment. Minimal depletion of *p*-coumaric acid was observed suggesting an issue in the conversion of *p*-coumaric acid to *p*-coumaroyl-CoA in the naringenin pathway (Figure 6.9B).

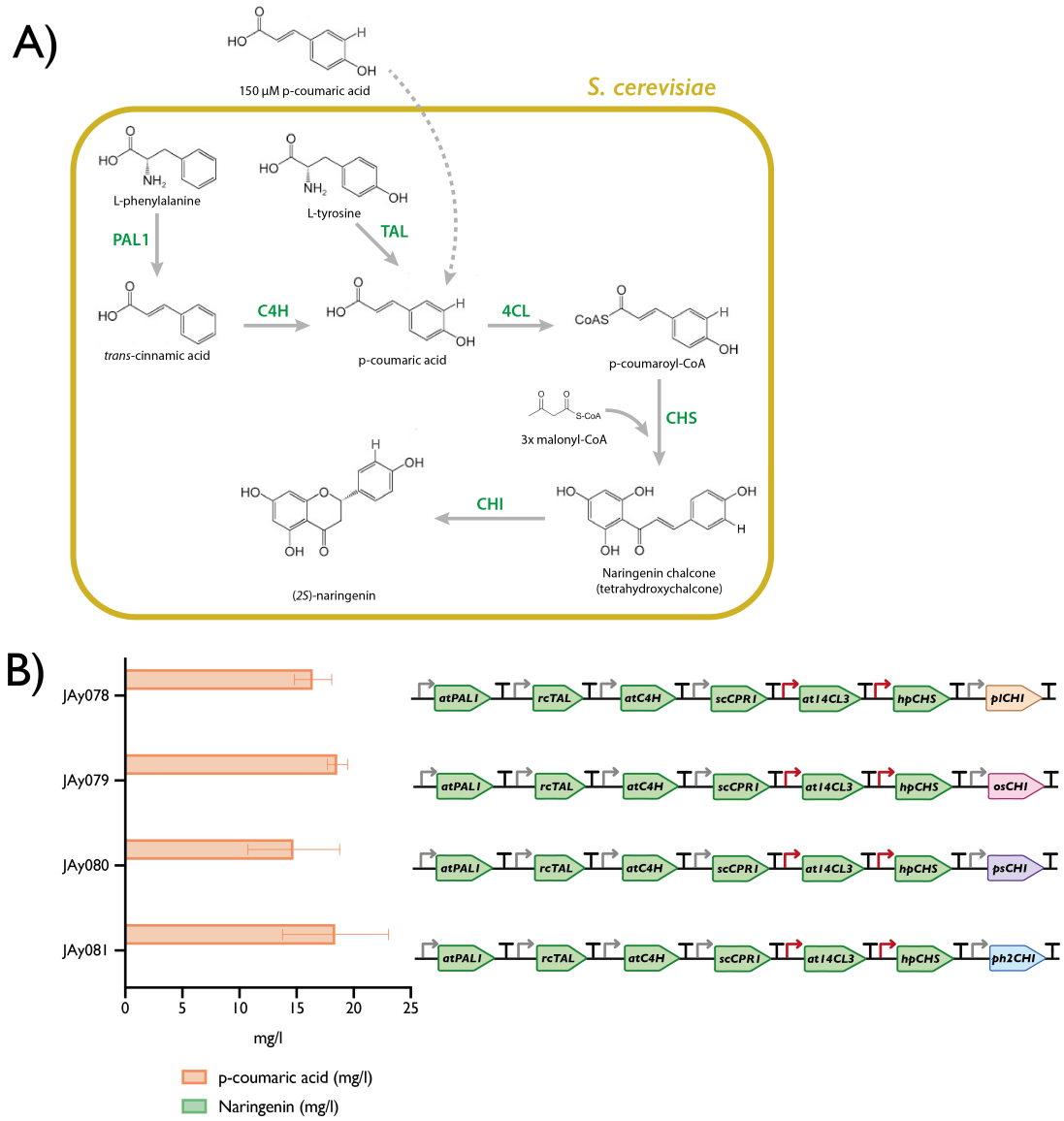


Figure 6.9

A) Feeding of naringenin-containing yeast strains with *p*-coumaric acid. B) Quantification of naringenin and *p*-coumaric acid production using the mass spectrometer. Production of naringenin by all 4 strains were not within detectable limits

6.2.7 Assembly of the Nar AT pathway

Meanwhile, the entire naringenin pathway was assembled using BSGs originating from *A. thaliana* with the exception of the *TAL* gene (*Rhodobacter capsulatus*). This pathway is a similar design to that assembled by Koopman *et al.* (Koopman et al., 2012). However, Koopman *et al.* separated the genes, with half hosted in a plasmid and the other half genomically integrated. In contrast, all naringenin BSGs in each strain were hosted on one plasmid (pYAC10).

This version of the naringenin pathway, Nar-AT, consist of 6 genes from *A. thaliana*). From the previous assemblies (NC2, NC3, NC4 and NC6), three BSGs (*CPR*, *CHS* and *CHI*) did not originate from *A. thaliana*. Therefore the rest of the BSGs (*PAL*, *C4H* and *4CL*) containing *A. thaliana* genes, along with *TAL* (originating from *R. capsulatus*), were reused for the assembly of Nar-AT.

The re-assembly of *CPR, CHS* and *CHI* POTs to contain genes originating from *A. thaliana* utilised the same set of promoters, terminators and POT vectors were used in the previous assembly of transcription units with *CPR*, at *CHS* and at *CHI* genes for NC2, NC3, NC4 and NC6 pathways. These new assemblies are highlighted in dark red in Table 6.4. Following the Golden Gate assembly, the reactions were transformed into *E. coli*. PCR verification of one clone for each assembly showed the presence of the same pattern as that predicted in the gel simulation (Figure 6.10). The sizes of the amplicon bands were of the same size as those predicted in Table 6.4.

Table 6.4

Assembly plan for POTs in the second round of naringenin pathway assembly, utilising biosynthetic enzymes from a single host, *A. thaliana* with exception to *TAL*. In green are the POTs reused from the round 1 assembly. In red are the new assemblies carried out for this round.

Name	Promoter	ORF	Gene length (bp)	Terminator	POT	Length of transcription unit (bp)	Predicted length of amplicon after PCR verification (bp)
N1	pRPS2	*atPAL*	2202	tACS1	POT2	2902	3034
N2	pACT1	*rcTAL*	1620	tCIT1	POT3	2320	2452
N3	pRPS3	*atC4H*	1542	tRPS13	POT1	2242	2374
N4-AT	pPGK1	*atCPR*	2103	tACS2	POT1	2803	2935
N5	pZ_3EV	*at1-4CL3*	1512	tFUM	POT1	2212	2344
N6-AT	pZ_3EV	*atCHS*	1212	tRPS3	POT2	1912	2044
C-AT	pTDH3	Z*atCHI*	654	tENO2	POT3	1354	1486

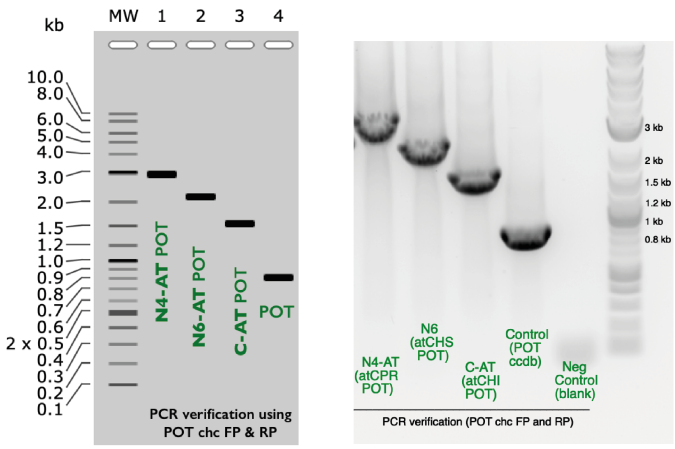


Figure 6.10

PCR verification of naringenin POTs in the second round of assembly

All three genes were verified for their expression in yeast by Western blot. Similar to the previous set of naringenin biosynthetic enzymes, the POT vectors were transformed into the SASy31 yeast strain, grown to an OD of 0.8 (exponential phase) and analysed for protein production. *atCHS*, which is driven by the pZ_3EV inducible promoter was grown in replicates, one induced and the other left uninduced. All three genes appeared to be expressed in yeast (Figure 6.11). Protein sizes were calculated for each *A. thaliana* BSG enzyme tested (Table 6.5). *atCPR*, *atCHS* and *atCHI* showed band sizes consistent with the predicted protein sizes, with *CHS* showing no expression in the uninduced condition (Figure 6.11). The total protein on the gel was imaged using the BioRad stain free technology (Appendix figure F.5), showed the loading of the same amount of protein lysate for each strain and any differences in the production of the FLAG-tagged protein are not due to the amount of total protein loaded.

Table 6.5

Predicted sizes of each naringenin *Arabidopsis thaliana* BSG enzyme (in kDa)

POT plasmid name	POT plasmid construct information	kDa
N4-AT	pPGK1-atCPR-tACS2 POT1	77.77
N6-AT	pZ_3EV-atCHS-tRPS3 POT2	44.12
C-AT	pTDH3-atCHI-tENO2 POT3	24.44

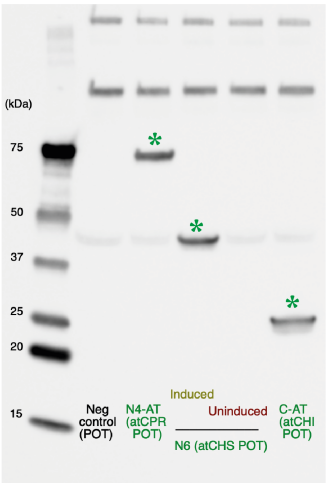


Figure 6.11

Western blot of naringenin BSG enzymes originating from *A. thaliana* expressed in yeast. A green asterisk (*) is used to label bands of interest on each lane.

Constructs N1 YFass1, N2 YFass2 and N4 YFass4 from the previous assembly were reused for this round of assembly (Table 6.6 - labelled in green). However, YFass 3 and YFass 5 assemblies were repeated to incorporate the genes from *A. thaliana* (Table 6.6 - labelled in red). Similar to previous YFass assemblies, BsaI was used to release the transcription unit assemblies from the POT vectors of N4-AT, N6-AT and C-AT. Two Golden Gate reactions were carried out, one for N3-AT YFass3 to host the *atCPR* transcription unit and N5-C-AT YFass5 to host *atCHS* and *atCHI* transcription units (Table 7). Assemblies were verified by colony PCR using primers specific for the *atCPR*, *atCHS* and *atCHI*, showing a similar band pattern to that observed in the gel simulation (Table of primers) (Figure 6.12).

Successfully verified N3-AT YFass3 and N5-C-AT YFass5 were digested with BsmBI and purified. Digested N3-AT YFass3 and N5-C-AT YFass5, along with previously digested NI YFass1 N2 YFass2 and N4 YFass4 were assembled into pYAC10 in the SASy31 yeast strain and selected on SC-His-Ura plates.

Table 6.6

Multi-tanscription unit (YFass) assembly plan for in the second round of naringenin pathway assembly, utilising biosynthetic enzymes from a single host, *A. thaliana* with exception to *TAL*. In green are the YFass assemblies reused from the round 1. In red are the new assemblies carried out for this round.

YFass assembly name	POT plasmid name	POT plasmid construct information	YFass vector
N1 YFass1	N1	pRPS2-*atPAL1*-tACS1 POT2	YFass 1
	N2	pACT1-*rcTAL*-tCIT1 POT3	
N2 YFass2	N3	pRPS3-*atC4H*-tRPS13 POT1	YFass 2
N4 YFass4	N5	pZ$_3$EV-*at1-4CL3*-tFUM1 POT1	YFass 4
N3-AT YFass3	N4-AT	pPGK1-*atCPR*-tACS2 POT1	YFass 3
N5-C-AT YFass5	N6-AT	pZ$_3$EV-*atCHS*-tRPS3 POT2	YFass 5
	C-AT	pTDH3-*atCHI*-tENO2 POT3	

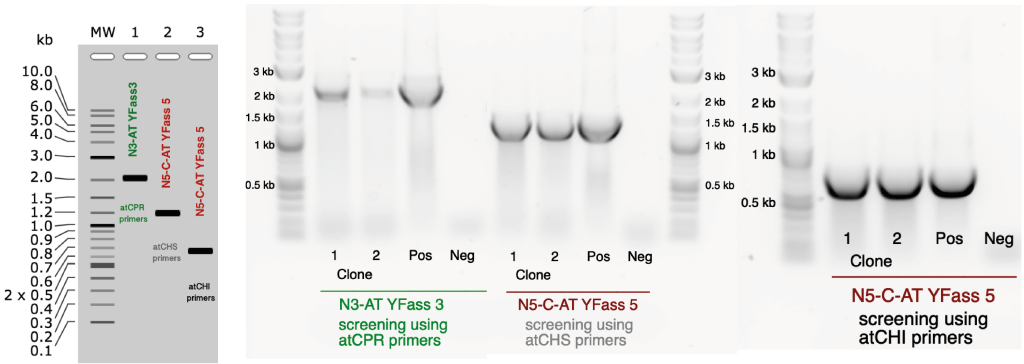


Figure 6.12

Gel simulation and agarose gel showing results for the PCR verification of each naringenin BSGs in each new YFass assemblies for this second round of assembly

JAy092 (Nar-AT yeast strain) was screened using primers specific for all seven naringenin BSGs. All BSGs were successfully amplified in one clone with the exception of the *atPAL* (Figure 6.13). Amplification of *atPAL* did not show any amplicon, however upon repeating the PCR reaction with a longer extension time, a faint band of the correct size was successfully amplified (Figure 6.13). This clone was subsequently screened for the production of naringenin by mass spectrometry.

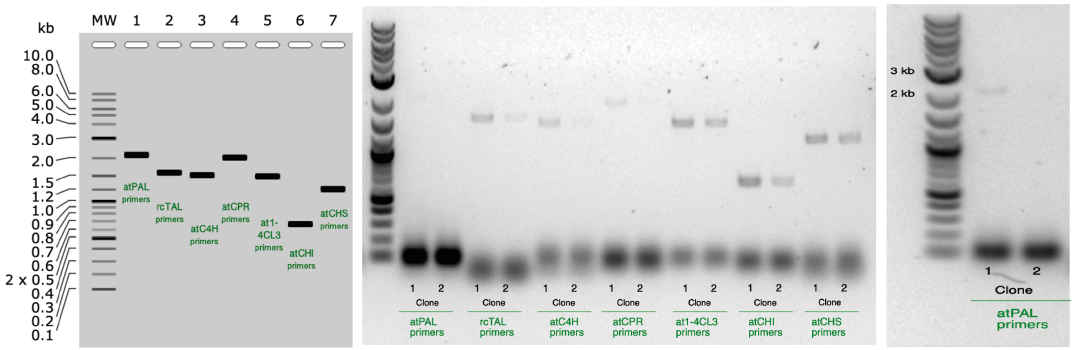


Figure 6.13

Gel simulation and agarose gel showing the colony PCR verification of all 7 naringenin BSGs in the JAy092 strain (Nar-AT pathway)

Similar to the previous measurements of naringenin production by strains containing different variants of the naringenin pathway, biological replicates of JAy092 were grown overnight in in SC-His-Ura medium with 1 μM β-estradiol and 150 μM *p*-coumaric acid. *p*-coumaric acid was spiked in the medium to determine whether a high supply of a key intermediate can result in naringenin production levels within detectable limits. Following 24 hours of growth, the yeast spent medium was analysed for the level of *p*-coumaric and naringenin production using the triple quadrupole mass spectrometer, with help on sample processing on the mass spectrometer and data analysis from Dr. Cunyu Yan. *p*-coumaric acid was detected in the spent media in similar amounts as in JAy078, JAy079, JAy080 and JAy081 (Figure 6.14). Similar to previous findings of naringenin production in JAy085, JAy086, JAy087 and JAy088 strains, naringenin was not within detectable limits, suggesting that the problem was not due to the CPR, CHS or CHI enzymes as at least 2 versions (6 versions for CHI) of each of these enzymes have been tested and no improvement was observed.

Previous verification of JAy079 and JAy081 without the introduction of purified *p*-coumaric acid to the media prior to strain growth showed the detection of *p*-coumaric by mass spectrometry (∼8-12 mg/l) suggesting that the first 4 enzymes in the pathway (PAL, TAL, C4H and CPR) are functional or that one of the routes (either via TAL or PAL, C4H and CPR) from an aromatic amino acid to *p*-coumaric acid is functional. This suggests that the problem in the pathway lies in the latter half of the pathway (consisting of 4CL, CHS and CHI enzymes). The Nar-AT pathway shows that the

utilisation of a different CHS and CHI variant did not result in naringenin production within detectable limits. This leads to the possibility of a defective 4CL enzyme.

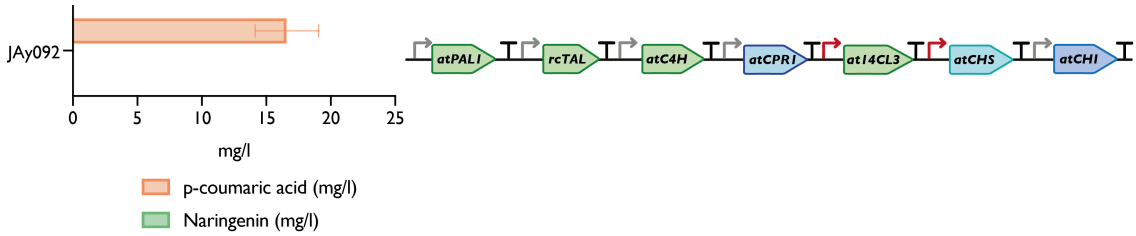


Figure 6.14

Quantification of naringenin and *p*-coumaric acid production in JAy092 (Nar-AT pathway) using the mass spectrometer. Production of naringenin by JAy092 was not within detectable limits

6.2.8 Analysis of the *4CL* enzyme

Previous work on the assembly of the naringenin pathway has shown that increasing the amount of enzyme available for the conversion of *p*-coumaric acid to *p*-coumaroyl-CoA resulted in significantly higher titres of naringenin (Watts et al., 2004). Analysis of *4CL* production by western blotting (Figure 6.5) in shows a lower expression of the enzyme compared to *CHS*, both of which are driven by the same promoter.

Further analysis of the 4CL enzyme was carried out, comparing the coding sequence for the previously tested *4CL* gene (*at1-4CL*) and other versions of this enzyme available in the SynNP library. Another *A. thaliana 4CL* gene was previously codon optimised by Dr. Almer van der Sloot from the Tyers lab and synthesised by the commercial vendor, Gen9, which encoded for a longer 4CL enzyme. I have previously domesticated this variant into the YeastFab format and, similar to other naringenin BSGs, incorporated a FLAG tag immediately downstream of the start codon (Appendix figure F.1). A comparison of the at1-4CL and at2-4CL protein sequences by Clustal Omega showed low sequence conservation in the first 40-50 amino acids of the sequence and high conservation for the rest of the protein. However, there appears to be a premature truncation in the C-terminus of the protein sequence of *at1-4CL*, with the last 69 amino acids missing compared to *at2-4CL* (Figure 6.15)

The at1-4CL protein sequence was also compared against the original protein sequence (NCBI reference number 849844.1 - labelled at4CL-NP_849844.1 in this

study) from which the *at1-4CL* sequence was based on, to identify the source of the premature truncation. The *at1-4CL* DNA sequence used in this study is the result of the codon optimisation of the original *A. thaliana* at4CL-NP_849844.1 DNA sequence. In the following protein sequence analysis, the translated amino acid sequences of these two 4CL DNA sequences (*at4CL-NP_849844.1* and *at1-4CL*) were used. The protein sequence analysis of at4CL-NP_849844.1 and at1-4CL showed the 69 amino acid truncation for both, suggesting that at4CL-NP_849844.1 was possibly a mis-annotated CDS or a non-functional splice variant. (Figure 6.15).

```
CLUSTAL O(1.2.4) multiple sequence alignment

at2-4CL                 MDYKDDDDKAPQEQAVSQVMEK----------QSNNNNSDVIFRSKLPDIYIPNHLSLHD     50
at4CL-NP_849844.1       --------mitaalhepqihkptdtsvvsddvlphspptprifrsklpdidipnhlplht     52
at1-4CL                 MDYKDDDDKITAALHEPQIHKPTDTSVVSDDVLPHSPPTPRIFRSKLPDIDIPNHLPLHT     60
                                      *: :             :.  :  ********* ***** **

at2-4CL                 YIFQNISEFATKPCLINGPTGHVYTYSDVHVISRQIAANFHKLGVNQNDVVMLLLPNCPE     110
at4CL-NP_849844.1       ycfeklssvsdkpclivgstgksytygethlicrrvasglyklgirkgdvimillqnsae     112
at1-4CL                 YCFEKLSSVSDKPCLIVGSTGKSYTYGETHLICRRVASGLYKLGIRKGDVIMILLQNSAE     120
                        * *:::*..: ***** * **: ***.:.*::*.*:::*:.::***:.:.**:*:** *. *

at2-4CL                 FVLSFLAASFRGATATAANPFFTPAEIAKQAKASNTKLIITEARYVDKIKPLQNDDGVVI     170
at4CL-NP_849844.1       fvfsfmgasmigavsttanpfytsqelykqlkssgakliithsqyvdklknlgenltli-     171
at1-4CL                 FVFSFMGASMIGAVSTTANPFYTSQELYKQLKSSGAKLIITHSQYVDKLKNLGENLTLI-     179
                        **:**:.**: **.:*:****:*  *: ** *:*.:*****.::*****:* * ::  ::

at2-4CL                 VCIDDNESVPIPEGCLRFTELTQSTTEASEVIDSVEISPDDVVALPYSSGTTGLPKGVML     230
at4CL-NP_849844.1       -----ttdeptpenclpfstlitd-detnpfqetvdiggddaaalpfssgttglpkgvvl     225
at1-4CL                 -----TTDEPTPENCLPFSTLITD-DETNPFQETVDIGGDDAAALPFSSGTTGLPKGVVL     233
                             . . * **.** *: *  .  *:. . ::*:*. **..***:***********:*

at2-4CL                 THKGLVTSVAQQVDGENPNLYFHSDDVILCVLPMFHIYALNSIMLCGLRVGAAILIMPKF     290
at4CL-NP_849844.1       thkslitsvaqqvdgdnpnlylksndvilcvlplfhiyslnsvllnslrsgatvllmhkf     285
at1-4CL                 THKSLITSVAQQVDGDNPNLYLKSNDVILCVLPLFHIYSLNSVLLNSLRSGATVLLMHKF     293
                        ***.*:*********:*****::*:********:****:***::* .** **::*:* **

at2-4CL                 EINLLLELIQRCKVTVAPMVPPIVLAIAKSSETEKYDLSSIRVVKSGAAPLGKELEDAVN     350
at4CL-NP_849844.1       eigalldliqrhrvtiaalvpplvialaknptvnsydlssvrfvlsgaaplgkelqdslr     345
at1-4CL                 EIGALLDLIQRHRVTIAALVPPLVIALAKNPTVNSYDLSSVRFVLSGAAPLGKELQDSLR     353
                        **. **:**** :**:* :***:*::*:**.  .:.*****:*.* **********:*::.

at2-4CL                 AKFPNAKLGQGYGMTEAGPVLAMSLGFAKEPFPVKSGACGTVVRNAEMKIVDPDTGDSLS     410
at4CL-NP_849844.1       rrlpqailgqgygmteagpvlsmslgfakepiptksgscgtvvrnaelkvvhletrlslg     405
at1-4CL                 RRLPQAILGQGYGMTEAGPVLSMSLGFAKEPIPTKSGSCGTVVRNAELKVVHLETRLSLG     413
                         ::*:* ***************:**********:*.***:*********:*:*. :*  **.

at2-4CL                 RNQPGEICIRGHQIMKGYLNNPAATAETIDKDGWLHTGDIGLIDDDDELFIVDRLKELIK     470
at4CL-NP_849844.1       ynqpgeicirgqqimkeylndpeatsatideegwlhtgdigyvdeddeifivdrlkevik     465
at1-4CL                 YNQPGEICIRGQQIMKEYLNDPEATSATIDEEGWLHTGDIGYVDEDDEIFIVDRLKEVIK     473
                         **********:**** ***:* **: ***::********* :*:***:********:**

at2-4CL                 YKGFQVAPAELEALLIGHPDITDVAVVAMKEEAAGEVPVAFVVKSKDSELSEDDVKQFVS     530
at4CL-NP_849844.1       fkgfqaspfifivlfifnkclhlkkiyimi------------------------------     495
at1-4CL                 FKGFQASPFIFIVLFIFNKCLHLKKIYIMI*------------------------------     503
                        :****.:*  : .*:* :  :    :  *

at2-4CL                 KQVVFYKRINKVFFTESIPKAPSGKILRKDLRAKLANGL* 569
at4CL-NP_849844.1       ---------------------------------------- 495
at1-4CL                 ---------------------------------------- 503
```

Figure 6.15

Protein sequence analysis of the *4CL* enzyme used in this study (at1-4CL) and its comparison with the original reference protein sequence from NCBI (NP_849844.1) and another *A. thaliana* enzyme in the SynNP library (at2-4CL)

6.2.9 Assembly of a truncated naringenin pathway

The pathway was truncated (removing the first segment of the pathway for the conversion of aromatic amino acids to phenylpropanoid precursors) to allow better control of the pathway during troubleshooting. With this, the pathway can be controlled via two inducers, 1) spiking with *p*-coumaric acid for testing the activity of *4CL* and 2) induction of *CHS* and *CHI* with β-estradiol.

In addition, Lyu *et al.* observed an additive effect in the naringenin yield due to different copy numbers of key enzymes in the pathway. They assembled different versions of the naringenin pathway to test the impact such copy number variation. With one copy of each gene, the yield was 5.2 mg/L. Naringenin yields increased to 8.8 mg/L when two copies of *CHI* were introduced, whereas two copies of *CHI* and *TAL* resulted in a doubling in naringenin yield (10.3 mg/L) compared to the single copy pathway. Doubling of all 4 genes in the pathway (CHI, *TAL*, *4CL* and *CHS*) resulted in a naringenin yield of 15.8 mg/L. From this observation, they suggested that the increase in *CHI* copy number compensated for the poor efficiency of spontaneous isomerisation of tetrahydroxychalcone (naringenin chalcone) at low pH (Lyu et al., 2017). Koopman *et al.* also increased the copy number of a key enzyme, chalcone synthase (CHS), due to the low catalytic activity of the enzyme (Koopman et al., 2012).

To test this, two copies of each gene were incorporated into the design of the truncated naringenin pathway. Two different *4CL* genes were tested for this assembly. One variant is another version of the *4CL* enzyme from *A. thaliana* and the other is a *4CL* gene from *Petroselinum crispum*. The *4CL* genes were placed under the control of pTDH3 while *haCHS* and *ph2CHI-6* were placed under the control of the pZ_3EV inducible promoter. With this design, *CHS* and *CHI* can be controlled by estradiol and the output of *4CL* is controlled by the presence or absence of *p*-coumaric acid. *at2-4CL* and *pc4CL* were both assembled into POT1-*ccd*b. All the POT assemblies in this round were carried out according to the design in Table 6.7. The assembled vectors were verified by PCR and showed a similar band pattern to that found on the gel simulation (Figure 6.16).

Table 6.7

Multi-tanscription unit (YFass) assembly plan for in the second round of naringenin pathway assembly, utilising biosynthetic enzymes from a single host, *A. thaliana* with exception to *TAL*. In green are the YFass assemblies reused from the round 1. In red are the new assemblies carried out for this round.

Name	Promoter	ORF	Gene length (bp)	Terminator	POT	Length of transcription unit (bp)	Predicted length of amplicon after PCR verification (bp)
N1-trunc POT	pTDH3	*at2-4CL*	1710	tACS2	POT1	2410	2645
N2-trunc POT	pZ_3EV	*haCHS*	1197	tFUM1	POT1	1897	2132
N3-trunc POT	pZ_3EV	*ph2CHI-6*	687	tENO2	POT1	1387	1622
N4-trunc POT	pTDH3	*pc-4CL*	1710	tACS2	POT1	2410	2645
N5-trunc POT	pZ_3EV	*haCHS*	1197	tACS1	POT2	1897	2132
N6-trunc POT	pZ_3EV	*ph2CHI-6*	687	tCIT1	POT3	1387	1622

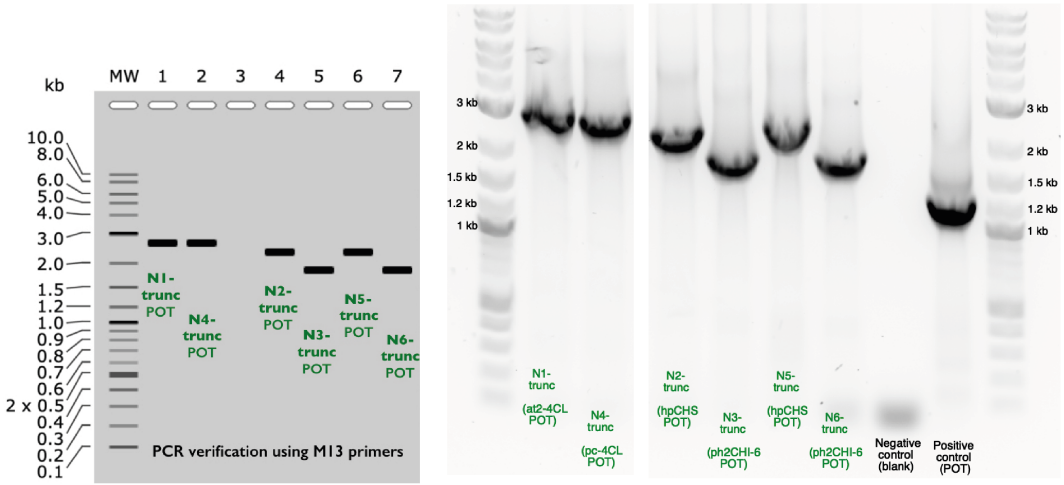


Figure 6.16

PCR verification of naringenin POTs in the truncated naringenin assembly

The resulting plasmids were transformed into yeast for the verification of protein production on a western blot. This was to determine which *4CL* enzyme variant appeared to be expressing a higher amount of protein when driven by the same promoter. at2- *4CL* was producing a similar amount of protein as *haCHS* and *ph2-CHI-6* (Figure 6.17). The total protein on the gel as shown in Appendix figure F.6 show similar amounts of protein lysates loaded for each strain. Therefore higher production of the FLAG-tagged protein is not due to the differences in total protein loaded on the gel. Screening of the other enzyme variants was not completed due to the limited duration and funding of the project.

Therefore, for the next round of assembly, two copies of *at2-4CL* were assembled along with 2 copies of *haCHS* and ph2-CHI-6. For this pathway design, 5 YFass assemblies were constructed according to the plan in Table 6.8.

YFass vectors were assembled by Golden Gate cloning and transformed into *E. coli*. These YFass vectors were subsequently screened by digestion verification. The agarose gel showing the products of the digestion showed a similar pattern to the digestion simulation made on SnapGene (Figure 6.18A).

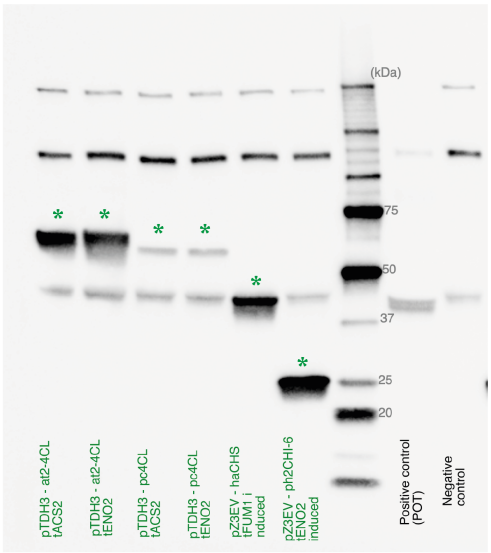


Figure 6.17

Western blot of naringenin *4CL* BSG enzymes. A green asterisk is used to label bands of interest on each lane

Table 6.8

Naringenin YFass assembly design for the truncated naringenin pathway. Labelled in dark green is the YFass using the same POT as N1-trunc YFass

YFass assembly name	Name	POT plasmid construct info	YFass vector
N1-trunc YFass 1	N1-trunc POT	pTDH3-at2-4CL-tACS2 POT1	YFass 1
N2-trunc YFass 2	N2-trunc POT	pZ$_3$EV-haCHS-tFUM1 POT1	YFass 2
N3-trunc YFass 3	N3-trunc POT	pZ$_3$EV-ph2CHI-6-tENO2 POT1	YFass 3
N4-trunc YFass 4	N1-trunc POT	pTDH3- at2- *4CL* -tACS2 POT1	YFass 4
N5-trunc YFass 5	N5-trunc POT	pZ$_3$EV-haCHS-tACS1 POT2	YFass 5
	N6-trunc POT	pZ$_3$EV-ph2CHI-6-tCIT1 POT3	

1.

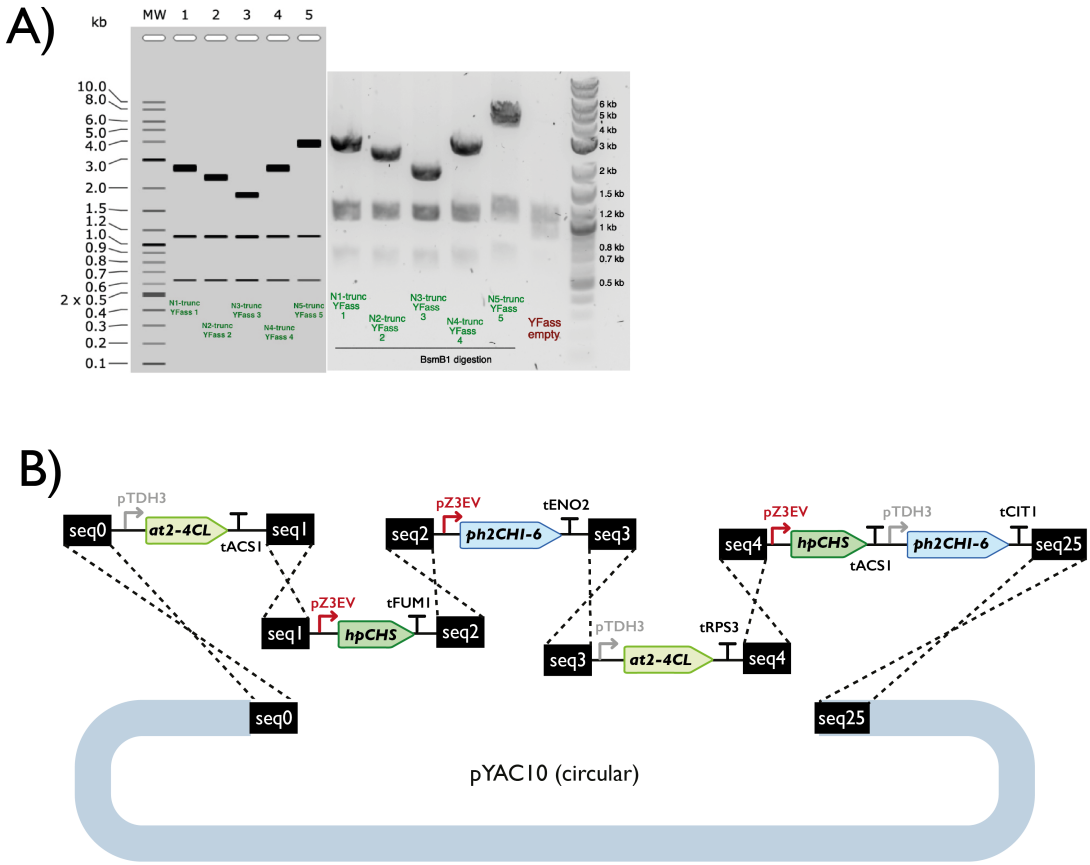


Figure 6.18

A) Comparison of the digestion simulation of truncated YFass vectors and the digestion verification results of truncated YFass vectors. B) Diagrammatic representation of the final truncated naringenin pathway containing two copies of *at2-4CL, haCHS, ph2CHI-6*

Digestion products were column purified and transformed into SASy31 along with a digested pYAC10 and selected on SC-His-Ura plates. Two clones were screened by colony PCR for the presence of all three naringenin BSGs (*at2-4CL*, *haCHS* and *ph2-CHI-6*) using gene specific primers (Table of primers). PCR results show the presence of all 3 genes in the pathway of the correct amplicon sizes (Figure 6.19). Clone 1 was grown and stored in a glycerol stock at -80°C under the strain ID of JAy093.

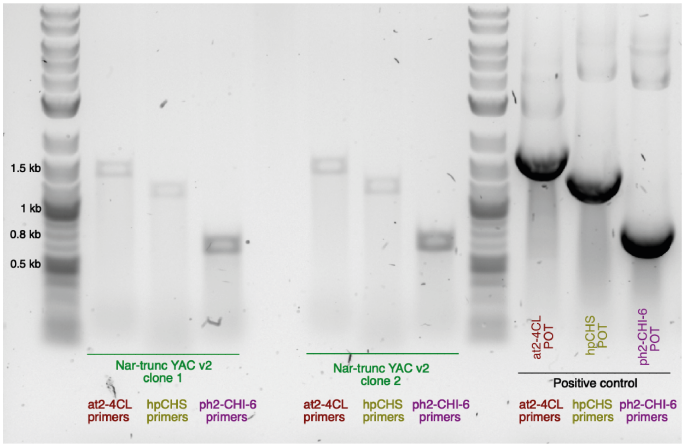


Figure 6.19

PCR verification of JAy093

Biological triplicates of JAy093 were grown in an overnight pre-culture in SC-His-Ura media at 30°C. The pre-culture was used to inoculate duplicates for each strain, one without and one with 1μM β-estradiol inducer for the activation of the pZ$_3$EV promoter. All strains were grown in fresh SC-His-Ura media with 1 mM of *p*-coumaric acid and analysed by mass spectrometry, with help on sample processing on the mass spectrometer and data analysis from Dr. Cunyu Yan. Modification of the pathway to include two copies of at2- *4CL* resulted in the depletion of *p*-coumaric acid and the production of naringenin (9 mg/L) (Figure 6.20). JAy093 was also sequence verified by Dr. Daniel Schindler and Marco Monti using the MinION (Oxford Nanopore), showing reads for all 6 genes (2 copies of *at2-4CL, haCHS and ph2CHI-6*) in a circular pYAC10 vector however due to low coverage of the sequencing reads, additional verification is required (Appendix figure F.7). We were not able to complete this due to time and funding restrictions.

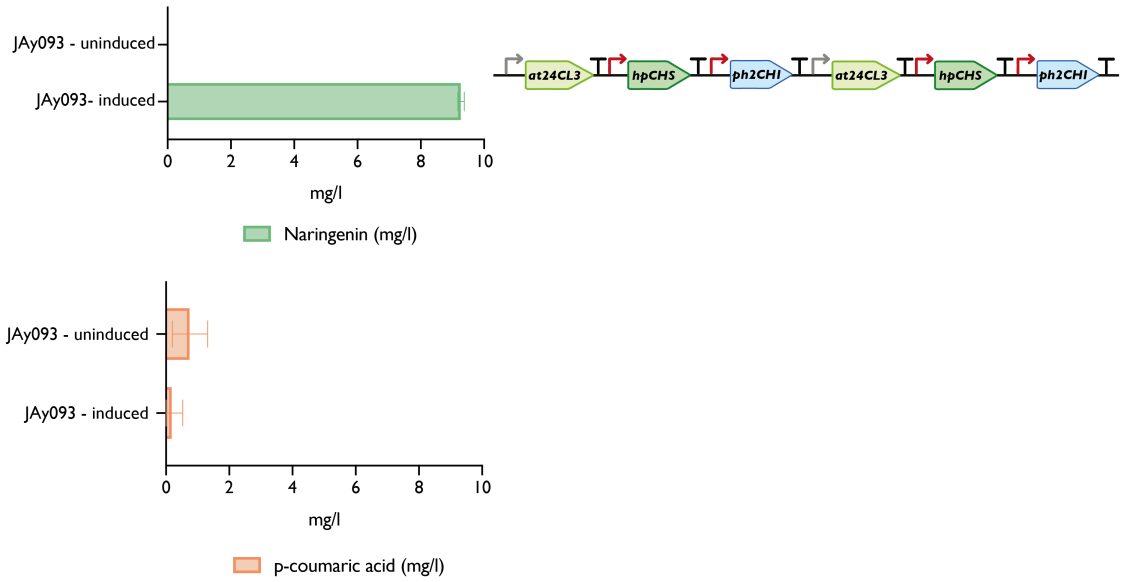


Figure 6.20

Quantification of naringenin and *p*-coumaric acid production using the mass spectrometer.

6.2.10 Establishing an assay for measuring naringenin production

In several publications showing naringenin pathway expression in both *E. coli* and yeast, researchers have found that 80-90% of detected naringenin was present in the spent media, requiring the development of an extracellular biosensor (Koopman et al., 2012; Trabelsi et al., 2018; Watts et al., 2004). An FdeR transcription factor-based biosensor was constructed and optimised by Trabelsi *et al.* for extracellular detection of pinocembrin and naringenin produced from *E. coli* strains (Trabelsi et al., 2018). The biosensor utilises a naringenin-sensitive FdeR protein which enables expression of RFP, upon FdeR binding to sequence motifs in the presence of the ligand, naringenin. Skjoedt et al. have translated this prokaryotic naringenin biosensor to yeast (Skjoedt et al., 2016) however my initial efforts to recapitulate their results were not successful (Data not shown). As a result, in the following work, the *E. coli* biosensor is optimised for the detection of naringenin in yeast spent media. Due to the acidity and presence of ethanol in yeast media, activity of the biosensor was screened in media of different pH and in fresh and spent yeast media. Optimal conditions obtained from initial screening tests are used for screening JAy093, an inducible producer strain previously

constructed, for the amount of naringenin produced. The biosensor is then used to detect the amount of naringenin produced in a small library containing variations in the enzyme used, gene copy number and background strain.

6.2.11 Initial screening of the naringenin biosensor

The naringenin (pACYC156) biosensor was provided as a gift from the Faulon laboratory and transformed into *E. coli* cells (Trabelsi et al., 2018). The biosensor strain was tested in three conditions, 1) without naringenin, 2) with 1 mM naringenin and 3) with 1 mM pinocembrin (Figure 6.21). After 24 hours of growth, the OD and RFP fluorescence (with an excitation of 545 and emission of 591) was measured on a plate reader in technical triplicates. Increasing amounts of fluorescence was observed when the concentration of naringenin added to the media was increased. When employing a Tukey's multiple comparison test, we clearly observe significantly higher levels of RFP when assessing uninduced compared to induction with 1 mM naringenin ($P < 0.0001$) as previously found by Siedler and colleages Siedler et al. (2014). In concordance with results from Trabelsi (2018) (Trabelsi et al., 2018) we not only observe significantly higher levels of RFP when assessing media only compared to 1 mM of pinocembrin ($P < 0.0001$), but also see significantly higher levels of RFP in the presence of 1mM of naringenin ($P = 0.0022$), albeit to a lesser extent than pinocembrin (Figure 6.21A).

Different concentrations of naringenin (10μM-250μM end concentration) were inoculated into the pACYC156 *E. coli* biosensor culture and grown over 24 hours in a plate reader. A slight decline was observed in the growth curve at stationary phase. However, comparing the growth curves of *E. coli* pACYC156, between the induction concentrations and uninduced condition, no difference in growth was observed (Figure 6.21B). This suggests that the factor contributing to the decline is not naringenin. The normalised fluorescence was calculated by normalising the fluorescence to the OD600 values. Differences in the fluorescence output (RFP) correlated to the amount of naringenin added to the media (Figure 6.21C). The endpoint normalised fluorescence value was used for generating a standard curve with an adjusted R2 value of 0.9963 (Figure 6.21D). Drawing from the principle of the Hill function, we can derive a sigmoidal curve that is able to model the predicted concentration when provided with calibrated data points from a standard curve (Trabelsi et al., 2018) (Materials and Methods section). Using this model, we can then interpolate from the standard curve

the expected naringenin concentration for any population where the RFP fluorescence and OD600 levels are known.

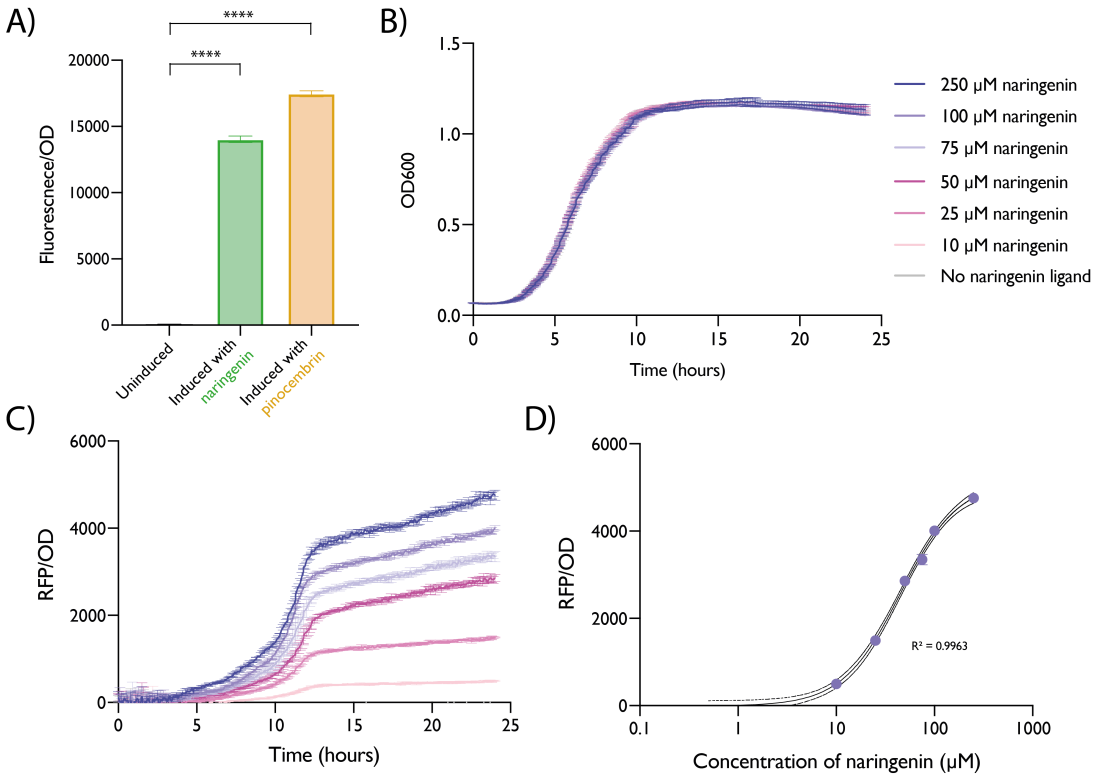


Figure 6.21

A) Screening for biosensor activation with the induction of naringenin and pinocembrin. B) Growth curves of the pACYC156 strain grown in different concentrations of naringenin. C) Screening of RFP/OD for the biosensor strain with different concentrations of naringenin. D) Analysis of the relationship between concentration of the inducer and the output fluorescence using nonlinear regression.

6.2.12 Testing for media conditions affecting the naringenin biosensor activity

Due to the differences in the pH of bacterial media compared to yeast media the activity of the pACYC156 biosensor was tested in LB at three media pH conditions (pH 4, pH 5.5 and pH 7), as low pH has been shown to compromise fluorescence as a result of chromophore protonation (Cranfill et al., 2016; Shinoda et al., 2018). Fluorescence

was observed only after 20 hours in *E. coli* cells grown in low pH (pH 4) with the addition of low concentrations of naringenin (10-25 μM). This could be explained either by the RFP reporter being compromised or the greatly reduced solubility of naringenin in solution at lower pHs (Lucas-Abellán et al., 2019). *E. coli* growth at pH 4 was also significantly compromised, with a naringenin dependent toxicity effect (Figure 6.22A). In contrast, no naringenin dependent toxicity effefcts seemed to be observed in *E. coli* growth in pH 5.5 and pH 7. No difference in growth was observed in pACYC156 *E. coli* strain growth at pH 5.5 compared to pH 7 (Figure 6.22B,C). At pH 5.5 and pH 7, minimal fluorescence difference was observed between strains grown in 100 μM and 75 μM whereas large differences in the fluorescence were observed between 10 μM and 25 μM of naringenin. The varying activity of the biosensor at different pH shows that further modifications will need to be carried out to optimise the biosensor for detecting naringenin in yeast spent media with a low pH.

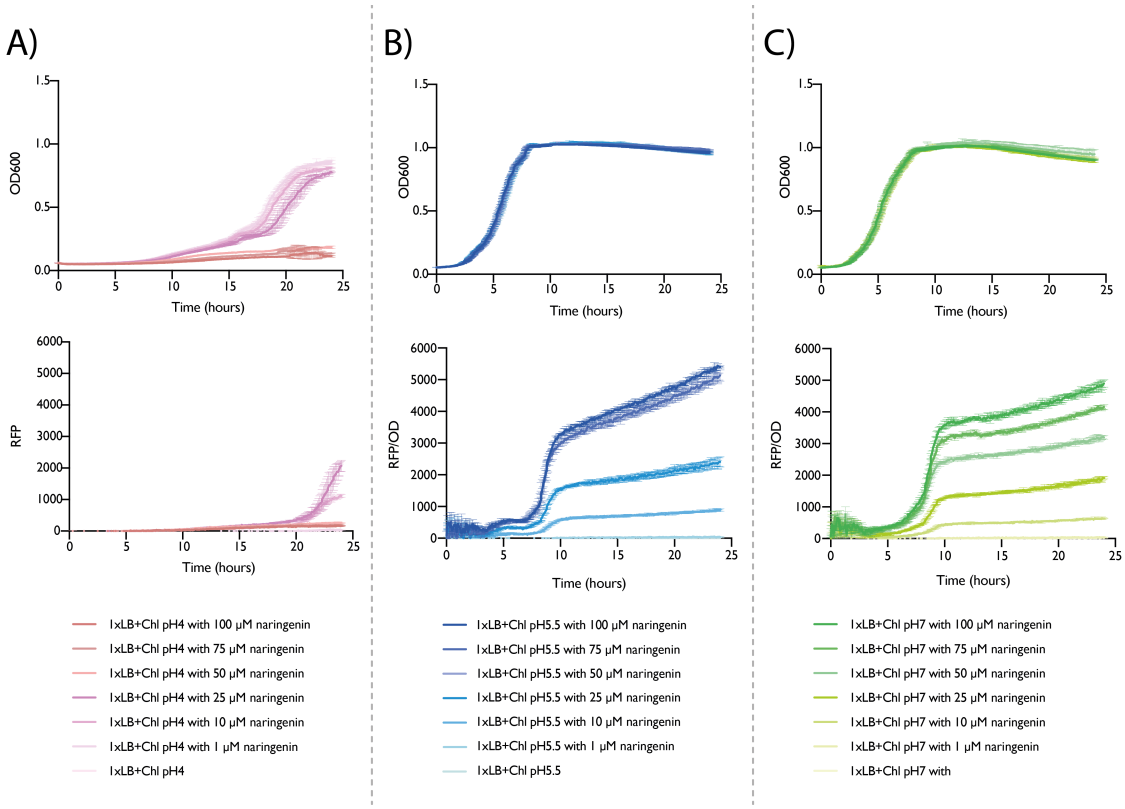


Figure 6.22

Screening of the biosensor in media with different pH

The biosensor was subsequently tested for growth in 75% minimal yeast media with 25% 4x LB, supplemented with different concentrations of naringenin. Growth of

pACYC156 *E. coli* strain in LB showed a similar growth and fluorescence profile as previously shown (Figure 6.23A). Growth in yeast minimal media resulted in a lower saturating OD (OD of 1.1 in normal LB conditions vs OD 0.8-0.9 in yeast minimal media) and ∼90% reduction in the fluorescence output for the pACYC156 biosensor in different naringenin concentrations (Figure 6.23B). Of note, despite the substantial drop in fluorescence observed, the output fluorescence still clearly correlates to the amount of naringenin present in the media.

To further explore the importance of pH, both LB only and 75% minimal yeast media with 25% 4x LB were phosphate buffered aiming to achieve ∼pH 7. Both media conditions were then tested for improvements in the growth or fluorescence output relative to unbuffered, upon exposure to different concentrations of naringenin. The decrease in OD600 observed in previous tests (Figure 6.21B, 6.22B and 6.22C) was not present in figure 24C, suggesting an improvement in growth phenotype when buffered. Buffering LB appears to have an effect on fluorescence as well, extending the lag time by 2-3 hours before the exponential increase in output signal (Figure 6.22C). No discernible lag time in growth was observed. Buffering the LB supplemented yeast media appears to have increased the saturating OD from 0.8 to 1.2 (buffered) (Figure 6.23D). However, it does not appear to have affected the peak fluorescence, which remains ∼90% lower (Figure 6.23D) than that observed in the buffered LB condition (Figure 6.23C).

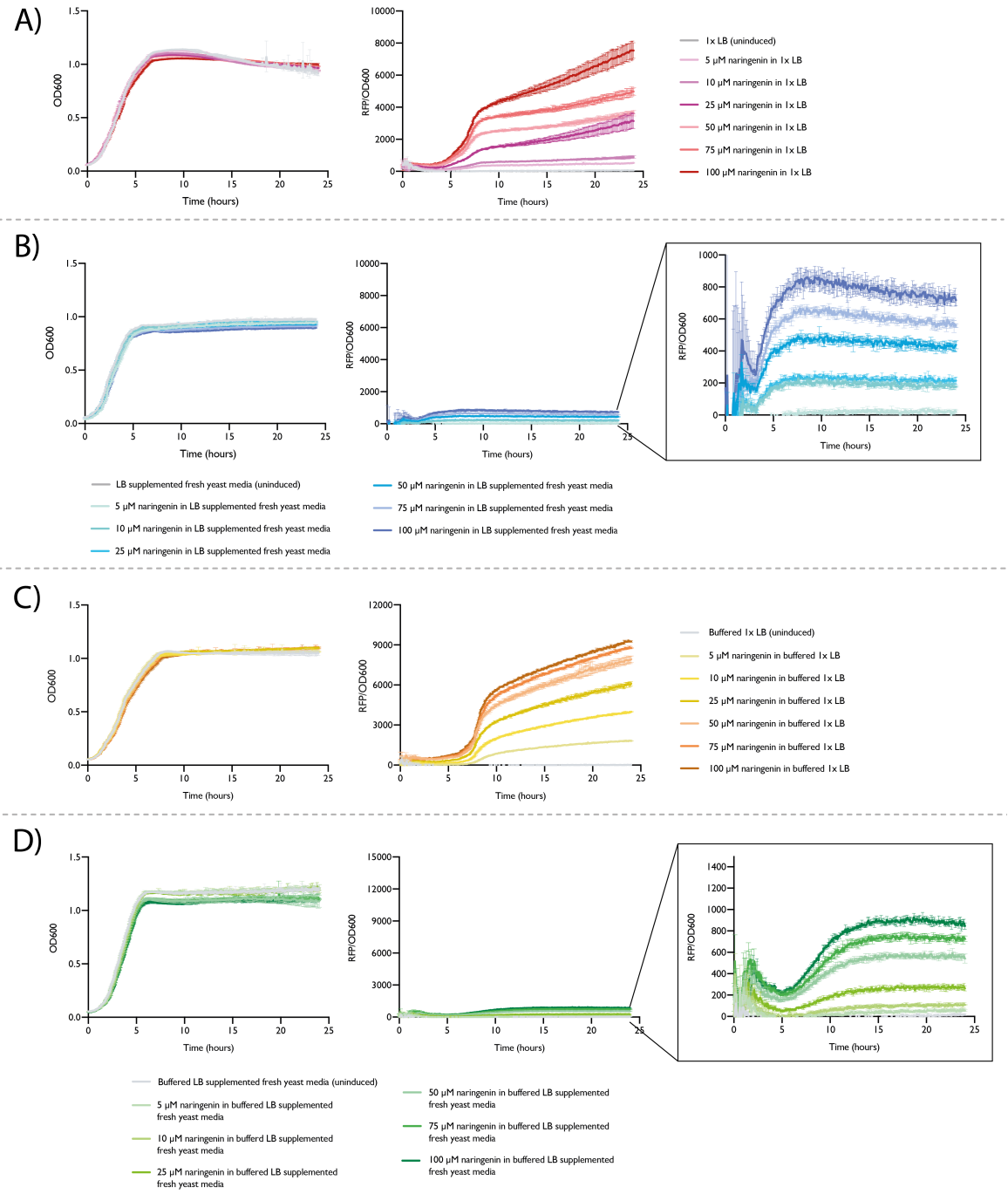


Figure 6.23

A) Screening of the biosensor in LB media, B) LB supplemented fresh yeast media, C) Buffered LB media, D) Buffered LB supplemented fresh yeast media

An initial comparison was carried out comparing the measured pH of fresh media and spent media. The spent medium consists of YAC10 grown in either synthetic complete media (SC) with ammonium sulphate (SC AMM) or SC with monosodium glutamate (SC MSG) for 24 hours then cells were pelleted by centrifugation to obtain the spent media. Yeast spent media of YAC10 grown in Sc MSG appears to have a slightly higher measured pH (∼4.5 pH) than Sc AMM spent media (∼4 pH) (Figure 6.24A). In all cases using fresh media the pH was around ∼6.5. However, in all cases using spent media, when one part of 4x LB was added to three parts media, the measured pH rose to pH 6.5. Screening of two yeast media formulations was due to previous findings by Lichten *et al.* showing that the primary source of fluorescence in yeast media was the ammonium salts (Lichten et al., 2014).

The naringenin biosensor was grown in these LB supplemented media over 24 hours, with 100 μM of naringenin introduced, to test its ability to detect the compound and produce an output RFP signal. For all spent media conditions (Figure 6.24B), there was a decline in OD in stationary phase suggesting an accumulation of a toxic or growth inhibiting compound. In contrast, growth in LB supplemented fresh media appears to have a higher saturating OD in both media formulations (Sc AMM and Sc MSG) than observed in the control wells with growth in buffered LB only. This suggests the utilisation of yeast media components for *E. coli* growth for a longer growth and higher saturation OD (Figure 6.24B).

Similar to previous findings, low fluorescence was observed in the naringenin biosensor grown in fresh yeast media. When grown in yeast spent media, the amount of fluorescent protein detected was at a similar level as the control (biosensor in LB medium only). A slightly lower amount was detected in Sc AMM spent media, however large error bars were observed in stationary phase, between hours 10 and 20. An increased abundance of RFP was detected in Sc MSG spent media at levels closely resembling those found for the control (Figure 6.24B).

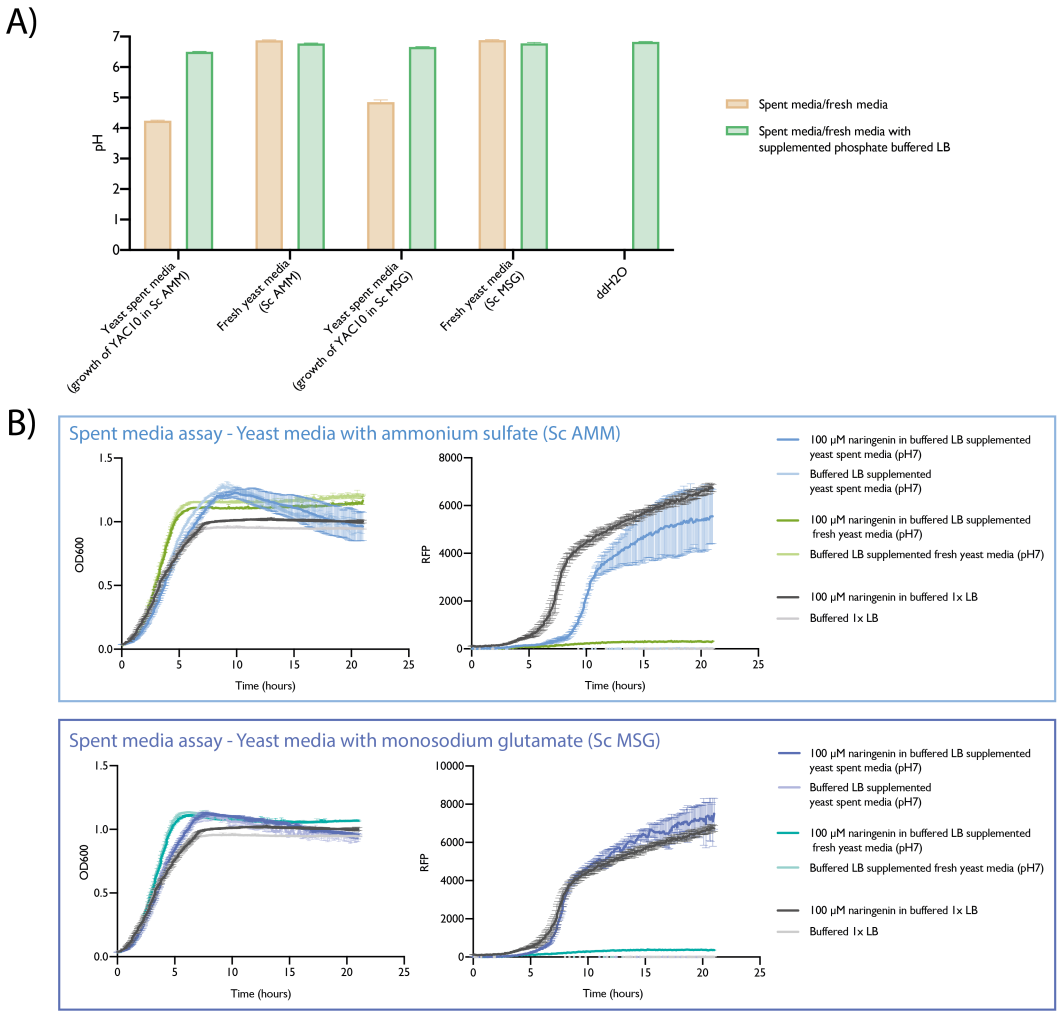


Figure 6.24

A) Measurement of media pH of yeast spent media and LB supplemented yeast spent media. B) Growth of the biosensor in yeast spent media with different nitrogen sources (ammonium sulfate and monosodium glutamate). In shades of blue are the results for the pACYC156 *E. coli* biosensor grown in buffered yeast spent media (containing ammonium sulfate). In shades of green are the results for the pACYC156 *E. coli* biosensor grown in buffered yeast spent media (containing monosodium glutamate). In shades of grey are the results for the pACYC156 *E. coli* biosensor grown in buffered LB media.

From the tests carried out, the optimal screening conditions appeared to be spent Sc media with MSG supplemented with buffered LB, which most closely reflected the original.

Using these conditions, the strain containing the biosensor was grown in the spent media of JAy093, a naringenin producing yeast strain. The yeast strain was grown in Sc minimal medium containing MSG for 24 hours then subsequently supplemented with buffered LB media. Two biological replicates of JAy093 were screened for naringenin production and compared to Sc MSG spiked with different amounts of naringenin to generate a standard curve. JAy093 was grown in three conditions; 1) medium only, 2) medium with 150 μM *p*-coumaric acid and 3) medium with 150 μM of *p*-coumaric acid and 1 μM of estradiol. JAy093 and a negative control (yeast strain containing an empty YAC) were tested in technical triplicates with biological replicates. Figure 25A shows the normalised fluorescence curves (RFP/OD) over time for each sample. In shades of green is the YAC10 yeast spent media with different concentrations of purified naringenin spiked in for generating a calibration curve. JAy093 spent media are labelled in shades of pink and purple. Only the samples which were induced with both *p*-coumaric acid and estradiol showed the production of RFP over time Figure 25A.

The end point reading (24 hours) of the standards was used to generate a standard curve (Figure 6.25B) for the interpolation of the mg/lLproduced in JAy093 samples based on their measured fluorescence. Similar to previous findings by Trabelsi *et al.*, a linear relationship was observed between the concentrations of (25 - 100 μM or 6.8 - 27.2 mg/L) naringenin (Trabelsi et al., 2018). The interpolated concentration in mg/L accounts for 75% of the naringenin concentration as 3 parts of the spent media assay consisted of the spent media and 1 volume consisted of the LB media. As such the interpolated amount was multiplied by 4/3 to calculate the expected concentration, which was plotted in Figure 25C. Welch's t test showed a significant difference between the interpolated amount between JAy093 with *p*-coumaric acid and JAy093 with *p*-coumaric induced with estradiol (p value = <0.0001). The absence of observable fluorescence in the *p*-coumaric acid only condition provided evidence that the biosensor is not sensitive to *p*-coumaric acid. The interpolated naringenin concentration for JAy093 with *p*-coumaric induced with estradiol ($\sim$9 mg/L) using the spent media assay showed similar results to the titres obtained by mass spectrometry in the previous chapter.

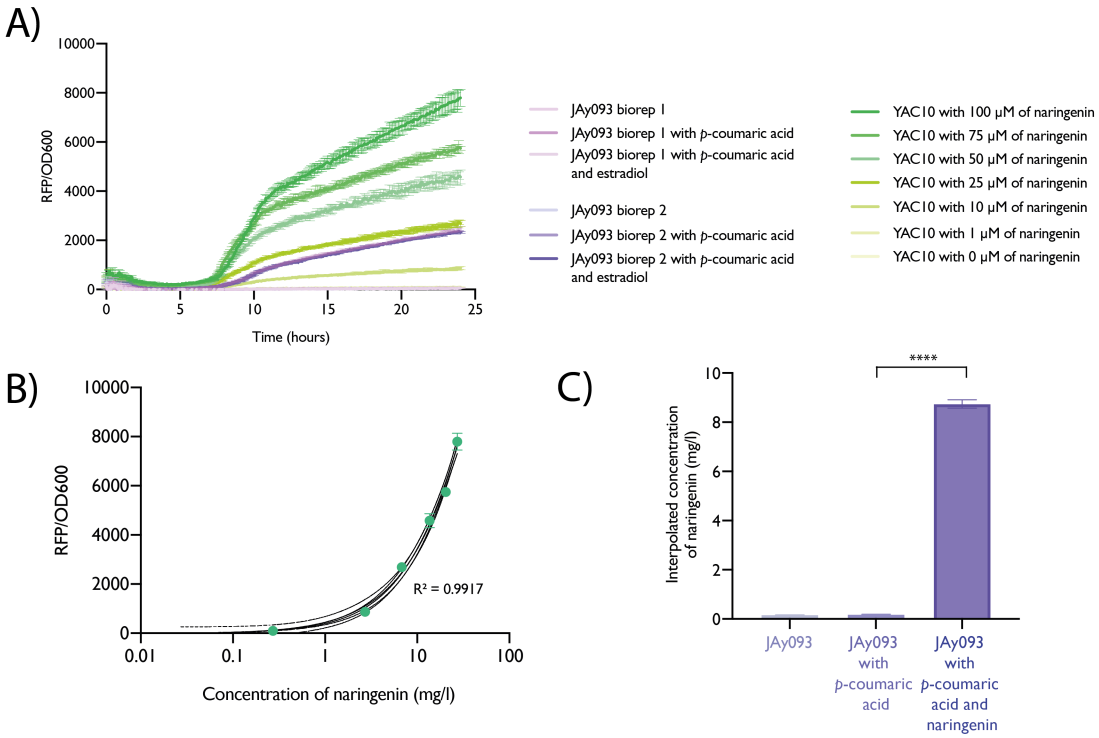


Figure 6.25

A) Growth of the *E. coli* biosensor in the yeast spent media of a naringenin producer strain (JAy093) over time. B) Generation of the standard curve for interpolating the concentration of naringenin produced by JAy093. C) Interpolated naringenin concentration of JAy093

The biosensor was subsequently used to screen for the production between a strain hosting the truncated naringenin pathway integrated in the HO locus on the yeast genome (JAy104) compared to its expression from a circular YAC (JAy093). This strain was constructed by using the previously digested YFass vectors for JAy093. In order to allow its integration into the HO locus, separate fragments containing homology regions to the HO locus were designed (Figure 6.26A). For the left homology arm, primers were designed to amplify the seq0 fragment from pYAC10 while incorporating an additional 60 bp homology to the HO locus. For the right homology arm, primers were designed to amplify the seq25 fragment and LEU2 marker from the pYAC10 vector while incorporating 60 bp homology to the HO locus. All 7 fragments were transformed into yeast and cells were selected in SC-LEU plates. JAy104 (HO integrated naringenin pathway in SASy31) was sequence verified, showing the integration of the fragment into ChrIV. JAy104, JAy093 and JAy096 (empty YAC) were screened for the

amount of naringenin produced using the spent media assay. JAy093 and JAy096 were grown in SC -His -Ura dropout media while JAy104 strains was grown in complete minimal media (with no amino acid dropped out). The three strains were grown in two conditions: uninduced medium containing 150 μM *p*-coumaric acid and in induced medium containing 150 μM *p*-coumaric acid 1 μM estradiol. Figure 6.26B shows the interpolated naringenin concentration corrected to account for the volume of spent media used in the assay (3 parts spent media, 1-part LB medium). When analysing the data using Tukey's multiple comparison test, significant difference in naringenin production was observed between the genomically integrated pathway (JAy104) compared to the plasmid hosted pathway (p value of 0.0001), with JAy093 showing a higher production ($\sim$9-12 mg/lL. Naringenin production differences between biological replicates of JAy104 were not statistically significant (p value of 0.0753). A baseline naringenin production ($\sim$1.2 - 1.8 mg/L) was observed in JAy104 strains when not induced by estradiol. Future work can seek to identify the factor contributing to the baseline fluorescence in the HO integrated pathway without estradiol induction. Comparison of the endpoint OD between plasmid containing strains and the genomically integrated pathway show a difference (OD of $\sim$0.3) in saturating OD (Figure 6.26C). This difference can be attributed to the difference in the media used (synthetic minimal -His -Ura dropout medium versus synthetic complete minimal medium) which can impact the OD600 reading. Another possible cause is the difference in yeast growth due to the growth media used. Further growth curve analysis could be carried out to verify this result however this was not completed due to time limitations.

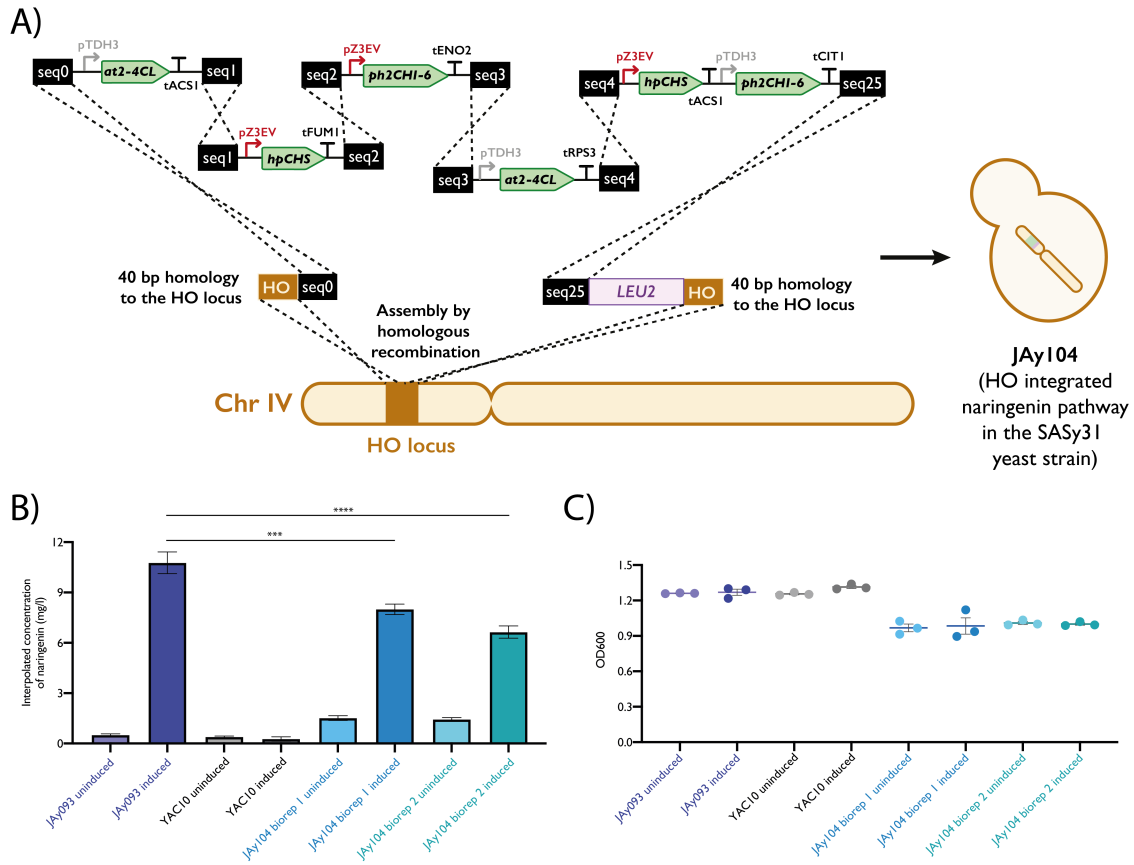


Figure 6.26

A) A diagrammatic representation of the strategy used for the integration of the truncated naringenin pathway into the HO locus in SASy31 yeast Chr IV. Strains are selected in Sc-Leu media to identify isolates containing the integrated pathway. The resulting strain was JAy104. B) The utilisation of the naringenin biosensor to measure naringenin yields of each producer strain, comparing a plasmid hosted (JAy093) versus a genomically integrated pathway (JAy104). Significant difference in naringenin production was observed between the genomically integrated pathway(JAy104) compared to the plasmid hosted pathway (p value of 0.0001, Tukey's multiple comparison test), with JAy093 showing a higher production (9-12 mg/L). C) Comparison of the endpoint OD between plasmid containing strains and the genomically integrated pathway

6.2.13 Screening for improvements in production through genomic rearrangements

Blount *et al.* and Liu *et al.* have previously shown several genomic rearrangements that resulted in improved compound production, penicillin and violacein for the former study and β-carotene and violacein for the latter study (Blount et al., 2018; Liu et al., 2018). Here, the JAy104 HO integrated naringenin strain was mated with the wild type haploid BY4742 strain or different synthetic yeast haploid strains generated through the Sc2.0 project (synII, synII/III, synII/III/VI/IXR) (Annaluru et al., 2014; Dymond et al., 2011; Mitchell et al., 2017; Shen et al., 2017). These strains were used for subsequent SCRaMbLE-mediated chromosomal rearrangements to explore whether genome engineering techniques can aid in strain optimisation for compound production. Initial screening was carried out using the biosensor to determine how yields of naringenin varied depending on the background strain (haploid vs. diploid, mated strain). All strains with the genomically integrated pathway showed a baseline fluorescence of (1.1 - 1.4 mg/L) in the condition without estradiol (Figure 6.27B).

This confirms results observed in Figure 6.26. No differences in yield were obtained between diploid strains, however large error bars were observed in JAy503 (JAy104 HO integrated pathway mated with BY4741). All measured naringenin concentrations in diploid strains appeared 2-fold less than the haploid JAy104. As previously observed, JAy093 showed the highest production ($\sim$ 10 mg/L). This difference can possibly be attributed to the use of different growth media (synthetic minimal -His -Ura dropout medium versus synthetic complete minimal medium) between the strains tested, confirming results previously obtained in Figure 26B. As for the strain growth, all the diploid strains analysed had an endpoint OD (OD 0.6 - 0.8) lower than JAy093 (OD of 1.2) (Figure 6.27C). JAy104 shows a higher OD than the haploid strains but lower than JAy093, as previously observed. No significant differences in growth were observed between induced and non-induced conditions across all strains.

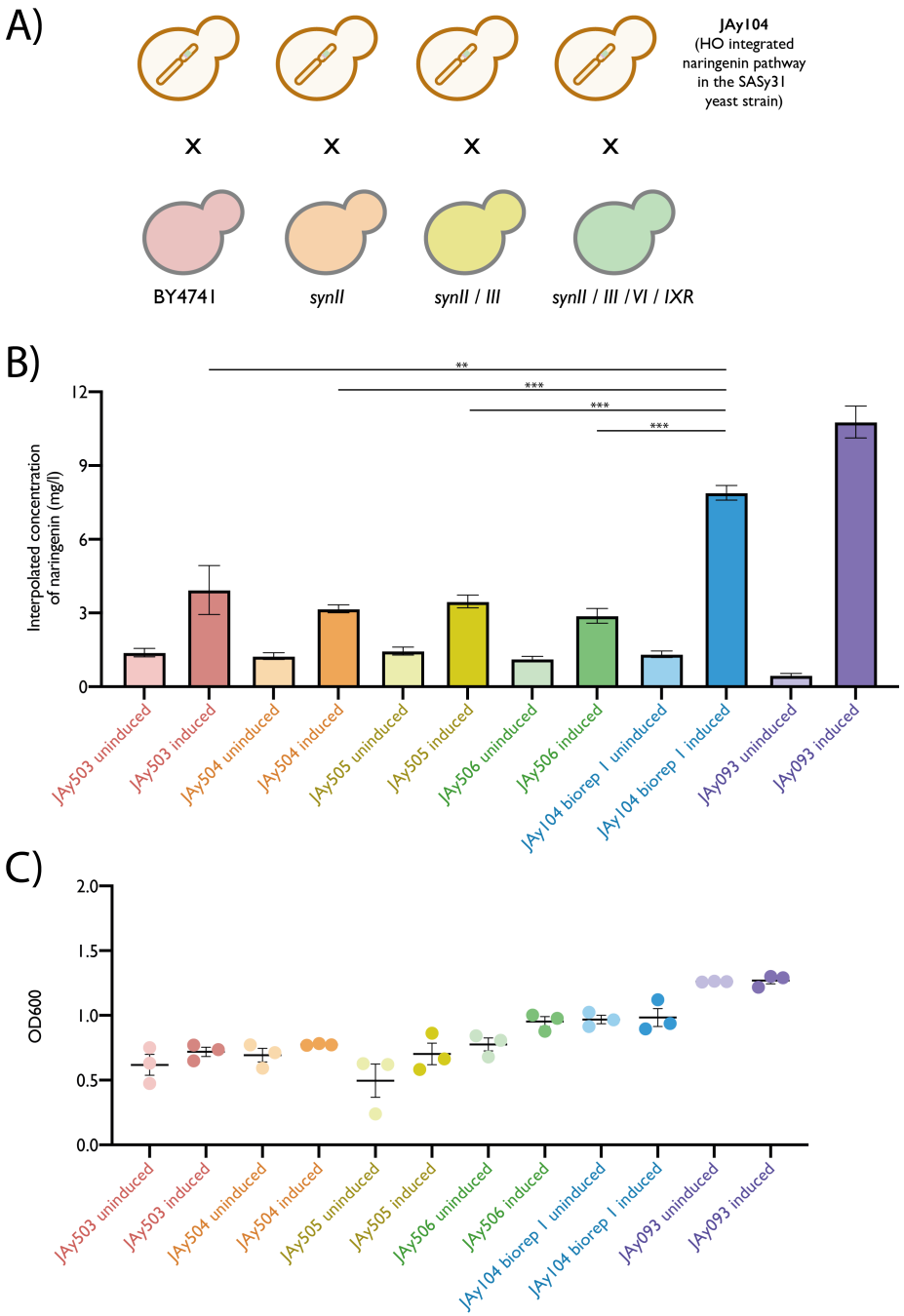


Figure 6.27

A) Diagrammatic representation of the strains mated for subsequent SCRaMbLE experiments for optimising naringenin titers in yeast. B) The utilisation of the *E. coli* spent media assay to measure naringenin yields of each producer/mated producer strain, comparing a haploid vs. a diploid strain. C) Comparison of endpoint OD measurements for each strain.

6.3 Chapter 6 discussion

This body of work demonstrates the utilisation of YeastFab for the rapid assembly and optimiation of the naringenin pathway. Initial assembly of the 7 gene naringenin pathway resulted in *p*-coumaric acid production however naringenin titerss were below detectable limits. Sequence analysis of the *4CL* enzyme revealed that the original variant (at4CL-NP_849844.1) for which at1-4CL was derived from possessed a truncation in the C-terminal. This C-terminal truncation of 4CL is likely responsible for the inactivity of the pathway (Figure 6.15). Using only the last 3 genes (*4CL, CHS* and *CHI* of the naringenin pathway, I have optimised the strain for the production of ∼8-12 mg/L of naringenin upon supplementation with *p*-coumaric acid after optimising the *4CL* enzyme. Future work will involve the assembly of the full pathway incorporating the optimised downstream pathway.

Several factors can be considered for incorporation into the pathway design to further increase yield. Koopman *et al.* observed protein-protein interaction between *CHS* and *CHI* derived from *A. thaliana* resulting in their decision to only utilise enzymes derived from the same organism . Along the same lines, synthetic protein scaffolds can be used to spatially localise biosynthetic enzyme genes, especially those known to interact, for improved production of a compound. This approach was used by Wang and colleagues for increasing resveratrol production in yeast (5-fold increase in titre) (Wang et al., 2011).

To date, there has not been any systematic effort to characterise the naringenin production in yeast of diverse pathways utilising different biosynthetic enzymes. Several publications utilise enzymes from a single organism while others use enzymes derived from multiple organisms. Within the SynNP library used in this work, multiple versions of biosynthetic enzyme genes were synthesised; two for *CPR*, six for *4CL*, two for *CHS* and seven for *CHI*. Using the YeastFab method, this library of enzymes can be used to generate over 168 assemblies, each with a different combination of biosynthetic enzyme genes. These strains can subsequently be screened using the spent media assay for high producers. A limitation of this library of enzymes is that it was only codon optimised using one set of parameters due to limited funding for DNA synthesis. Codon optimisation can impact the secondary structure of transcripts, impacting stability motifs (Dvir et al., 2013), playing a central role in translation rate. As such, future work can incorporate variants with different codon optimisation parameters along with the screening of more enzyme variants for improved activity. With the

number of combinations of enzymes along with different codon optimisation parameters, the incorporation of automation tools can be used to enable the rapid assembly and screening of naringenin pathway combinations.

Another method of improving yields is by engineering the endogenous metabolism for improved precursor supply. The availability of tyrosine and phenylalanine can be increased through the modification of the aromatic amino acid pathway. Many studies have been carried out to disrupt the feedback inhibition of DAHP synthase and chorismate mutase, allowing the high accumulation of tyrosine and phenylalanine in the cell (Koopman et al., 2012; Lyu et al., 2017). Further modifications to the yeast metabolism have been made to prevent the associated accumulation of phenylethanol which otherwise results in a decrease in growth (Koopman et al., 2012).

Promoter, RBS and terminator engineering are other options for optimising yields. Many groups have utilised a DNA assembly toolkit with a library of domesticated promoter parts to optimise yields (Wu et al., 2014). In particular Mitchell *et al.* have used their assembly pipeline, VEGAS (Versatile genetic assembly system), for the optimisation of beta carotene production in yeast (Mitchell et al., 2015). The success of such approaches demonstrates the value of combinatorial assembly of not only the coding sequence of different enzyme variants but further the genetic elements associated with each gene within a pathway. Golden Gate assembly methods, such as YeastFab offers a similar advantage – allowing the combinatorial assembly of 'parts' (promoters, ORFs or terminators) to generate a library of genetically diverse variants of the pathway. This capability was demonstrated in chapter 5 for the assembly of the viomycin pathway however due to time constraints, it was not carried out for the naringenin pathway.

However, a consequence of increasing the dimensions of exploration is an associated combinatorial explosion, whereby increasingly large numbers of assemblies would be needed to robustly sample the design space.

Advances in automation tools and the increasing development and optimisation of protocols have enabled the construction and screening of larger libraries of pathway variants. Kanigowska *et al.* optimised several protocols (PCR, Golden gate assembly and Gibson assembly) down to nanolitre scale (Kanigowska et al., 2016). The establishment of these protocols and availability of automation tools in the Cai laboratory enabled the scaling up of DNA assembly reactions to hundreds of assemblies allowing us to probe the contribution of multiple factors such as the promoter or enzyme choice

and the location of integration in the genome. However due to time limitations, I was unable to incorporate these technologies to the naringenin pathway optimisation.

Over the past decade, we have seen a surge in tools developed to alter the genome either mutating single or a handful of genes, or introducing genome wide mutations and rearrangements. One example, the SCRaMbLE system (Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution) relies on the use of synthetic yeast strains. The synthetic yeast genome project is an international collaboration aiming to resynthesise and reconstruct all 16 chromosomes in *S. cerevisiae* but incorporating several design principles. Some of these include the removal of introns, the recoding of all stop codons to TAA and the incorporation of loxP-sym (symmetrical LoxP) sites downstream of every non-essential gene in specific chromosomes (Pretorius and Boeke, 2018). This incorporation of loxP-sym sites enables inducible chromosomal rearrangements (inversion, duplication, deletion and translocation) by Cre-Estradiol-binding domain in the presence of estadiol. This approach has already been used to optimise the production for penicillin, β-carotene and violacein (Blount et al., 2018; Liu et al., 2018). Future work can explore the application of SCRaMbLE for the optimisation of naringenin yields. Whilst SCRaMbLE is a great technique for generating multiple highly diverse strains with ranges of activity, subsequent deconvolution and identification of high production strains can be challenging when production of a natural product compound does not directly promote survival (ie. confers improved survival in a stress condition or a nutrient limiting media) or provide a readily visualised output (ie. colourimetric output as in carotene or violacein producing yeast strains). When SCRaMbLE is applied for the production of a natural product, such as naringenin with no visual output (ie. coloured compound) or does not confer any benefit to the cell, it can be difficult to sift for high producers. It is also likely that a large population of strains resulting from genome rearrangement events mediated by SCRaMbLE may not result in high titers of the natural product as the compound incurs a metabolic burden to produce and at the same time, does not confer any advantages to the cell. As such one key bottleneck that emerges is the generation of scalable and ideally generalisable methods for selection of high producing strains for broader classes of molecules.

One approach to open this bottleneck is through the validation and use of biosensors. When appropriately optimised, biosensors enable the effective transduction of information, converting the concentration of a compound of interest into a visual output, such as fluorescence. A wealth of advances in research, protocols and hardware

allow the fluorescence to be used for sorting of high production strains and subsequent characterisation, with the potential for iterative cycles of optimisation.

A key innovation within the preceding work was the application of a biosensor to enable a fluorescence read-out proportional to naringenin abundance at a low cost relative to analytical techniques such as mass spectrometry. Not only does this enable subsequent cost-effective and high-throughput screening for naringenin but it can be applied to for the screeninig of other high value natural products, due to the prevalence of small molecule responsive transcription factors within nature (Koch et al., 2019; Voyvodic et al., 2019). This biosensor-based screening is also amenable to further expansion, allowing the screening for high producers in larger populations of genetically diverse strains modified by modern genome engineering approaches such as SCRaMbLE (Blount et al., 2018; Liu et al., 2018), eMAGE (Barbieri et al., 2017) and CHAnGE (Bao et al., 2018).

The establishment of this biosensor-based spent media assay should allow us to screen hundreds of strains from both pathway and genomic engineering efforts, each having different combinations of promoters, enzyme variants and terminators and pathway regulation as well as genomic rearrangements, and thus identify yeast strains with improved precursor flux or naringenin production efficiency in the yeast cell.

In the second half of this chapter, I discuss the optimisation of a spent media assay for the detection of naringenin, which is often exported to the extracellular milieu. In a collaboration with the Suess laboratory in Darmstadt, initial publication of a naringenin-sensitive riboswitch showed the requirement for a high amount of the compound for switching to occur resulting in the downregulation of the reporter gene (unpublished data). Therefore, we decided to optimise the FdeR based naringenin biosensor described by Trabelsi *et al.* for extracellular detection of naringenin produced from yeast strains (Trabelsi et al., 2018). Many factors that affect the activity of the biosensor were identified, for example growth of the *E. coli* was highly dependent on media pH, with optimal growth and functionality of the biosensor seen in LB media at a pH of 7. Initial screening of the biosensor in fresh yeast media show *E. coli* growth with a compromised activity (90% decrease in fluorescence activity). However, screening of yeast spent media and its buffering to a pH of 6.5 show a drastic improvement in biosensor activity to similar levels found in the positive control (LB media). An improvement in biosensor activity was also observed in the replacement of the yeast nitrogen source from ammonium sulphate to monosodium glutamate. Many factors

have not been thoroughly explored for their contribution to the biosensor system due to insufficient time. The biosensor optimisation carried out in this chapter utilised a one-factor-at-a-time method for understanding the different factors contributing to the system. The limitation of this approach is that it does not take into consideration the dependence of different factors one one another may significantly affect the system. Factors explored in the work above include pH of the yeast media prior to yeast growth, pH of the media prior to *E. coli* growth, nitrogen source of yeast media, phosphate buffering of the yeast media and the ratio of yeast spent media to LB media. Optimisation of each of the factors resulted in a spent media assay that could interpolate the amount of naringenin produced in JAy093 to similar levels observed in mass spectrometry measurements. However, several known factors still remain unaccounted for; impact of each yeast media component, the salt concentration of the buffered LB media, amount of *p*-coumaric acid added to the media, the impact of ammonium sulphate on *E. coli* growth compared to monosodium glutamate and length of culture time for both *E. coli* and yeast. These factors may affect the activity of the biosensor and if not controlled for, could lead to results that are not reproducible. As a result, future work can explore the use of Design of Experiments (DoE) to simultaneously investigate multiple factors that influence the biosensor-based screening method.

One potential approach to circumvent the issues associated with growth conditions is to increase the dilution of the yeast spent media in *E. coli*. Instead of the yeast spent media forming 75% of the media volume, this can be reduced to only constitute 10%. The fluorescence produced by the strain can used to back calculate the total concentration. This would also allow better quantification of higher naringenin production by a yeast strain, as current upper limit is $\sim$100μM or 27.225 mg/L). Another alternative would be to optimise the biosensor in a cell-free system, providing a controlled system that is not reliant on specific media conditions. In theory this method appears a promising option as it eliminates the impact of *E.coli* toxicity and import efficiency of naringenin on biosensor functionality. This strategy was utilised by Abatemarco *et al.* for identifying improved small molecule production (tyrosine, tryptophan, phenylalanine and methionine) in a library of *S. cerevisiae* strains using Spinach-based aptamer sensors co-encapsulated in a droplet using the protocol they developed called RAPID (RNA-aptamers-in-droplets) (Abatemarco et al., 2017). However extensive optimisation would likely be required to adapt this strategy for naringenin detection.

Chapter 7

Discussion

The preceding work has focussed on initially building upon the latest synthetic biology tools, seeking to expand upon the available pool of promoter parts before performing temporal analysis of promoter activity. This initial work expanded upon previous work that took single or in one case 4 time point measurements to reflect the activity of a single promoter. We show that whilst the normalised fluorescence value for a promoter at any single time point is partially reflective of the activity over time, there is up to 4-fold change in activity over 20 hours of culturing, with changing in the expression profile varying across a spectrum, with some of the most distinct divergences seen between promoters such as pRPS3 which shows peak activity during exponential phase (∼200 mins) followed by continual decline during culturing and pHSP26 which shows lowest activity during exponential phase (∼300 mins) before rising sharply, peaking in stationary phase (∼800 mins). Such nuances become impossible to infer from sampling 4 time points, especially when considering the varied profiles and peaking activity times observed across the library of promoters screened in chapter 4. As such when considering these promoters for metabolic engineering such profiles may be taken into account alongside the relative activity, especially for multi-step pathways where accumulation of intermediates may occur over several hours and when downstream enzymes display toxicity at high concentrations.

Whilst this work has helped to expand the available information associated with a subset of the library, in the future analysing more of the YeastFab promoter library will both increase the value of the resource for other researchers but may more specifically help identify other promoters with similar profiles to pHSP26 but differing expression

levels. One consideration and an area of expanding research is looking into the roles that terminator sequences and 3'UTRs play gene expression dynamics. Of note the preceding work has only analysed different promoter activity when coupled with the same terminator sequence so as to minimise the number of variables in flux. It is likely that some terminator sequences would change the fluorescence read-out across a range of promoters and it is possible that some 5'UTR/3'UTR combinations may show some measure of dependency, affecting measurements.

In chapter 4 we also see the application of an 'OFF' riboswitch for ligand dependent activity. We see that the riboswitch is able to function when embedded upstream of a different fluorophore cassette, implying portability. However of note we did observe a diminished switching factor as compared to the original publication, suggesting that further optimisation may be needed in some cases. When considered more broadly, an 'OFF' riboswitch system such as this could be very useful for regulating the production of especially toxic intermediary products if a specific riboswitch was embedded upstream of the enzyme responsible for production of the intermediary.

In chapter 5 the capabilities of the YeastFab system are tested, through the generation of the 26 gene viomycin pathway. We see not only successful assembly of the full length pathway but also show successful combinatorial assembly of the diversified pathway with each gene being driven by different promoters of varying activity. Due to the high number of screens generated we sought to apply a high-throughput screening strategy for the detection of strains producing an antimycobacterial compound. We identified a strain that showed visible antimycobacterial activity and this was subsequently screened by mass spectrometry. The strain did not produce detectable levels of viomycin, however high levels of palmitoleic acid, a known anti-microbial was produced. This emphasises the need for screening methodologies to enable selection based upon the closest possible characteristics to the compound of interest as is achievable to minimise the chance of 'cheaters', that mimic the desired phenotype, emerging.

In chapter 6 we again explore the use of the YeastFab system for expression of an antimicrobial compound, the flavanoid naringenin. Initial testing of the Naringenin biosensor (data not shown) in yeast did not show discernible activity. However we were able to demonstrate the use of the biosensor expressed in E.coli for detection of naringenin, when cultured with the spent media from naringenin producing yeast strains. In many ways this is more desirable, as it has previously been shown that naringenin is exported from yeast and as such endogenous biosensors may under-

represent production and bias screens towards strains that have reduced export of the compound, despite export being industrially useful. Such a biosensor screen should also be amenable to co-culturing within droplets (Baret et al., 2009; Abatemarco et al., 2017)which could facilitate much higher throughput to truly capitalise on the wealth of diversity achievable through combinations of pathway engineering and genome engineering. Given the abundance of naturally found riboswitches alongside developments in technologies such as SELEX, riboswitches should help provide a path forwards for very high-throughput initial screening for much broader classes of compounds. This in turn could alleviate a key bottleneck in the screening stages and help ensure that we don't continue instinctually reach for artemisinin as the success story for synthetic biology approaches in future years. Of note however these approaches are less suitable for compounds which are intrinsically less able to interact with nucleic acids, necessitating alternative strategies.

In the preceding work we have taken from start to finish the process of identification and domestication of biological parts, their characterisation and application for generation of diverse pathways for production of different classes of compound. We proceed through towards the successful utilisation of scalable screening approaches using a compound specific biosensor. Through pipelines such as this, it should become quicker, simpler and cheaper to generate and screen diverse pathway and strains for activity before isolating high producers for industrial production or further cycles of diversification.

Appendix A

Primers used in this study

A.1 Primers

Table A.1 Table of Primers

Primer name	**Primer Sequence**	**Purpose**
HcKan chc_F	GATCCTTTGATTTTCTACCG	Primers to sequence verify the HcKan insert
HcKan chc_R	CTCGATAACTCAAAAAATACG	Primers to sequence verify the HcKan insert
POT scr F	CAAAAGCTGGAGCTCCACCG	Primers to sequence verify the POT insert
POT scr R	TAGGGCGAATTGGGTACCGG	Primers to sequence verify the POT insert
M13 FP	GTAAAACGACGGCCAGTGAG	Primers to sequence verify the POT insert. Also used for the verification of inserts into the TOPO blunt II sub cloning vector.

M13 RP	CACAGGAAACAG CTATGACCATG	Primers to sequence verify the POT insert. Also used for the verification of inserts into the TOPO blunt II sub cloning vector.
POT1-ccdb (-BsaI) Gibson FP	GCTCCACCGCGGT GGCGGCCgcgaattcc ccgggtctagaGGTCTC TACCTGGCTAGAG ACGggatccggcttactaa aagccag	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT1 vector (by Gibson assembly)
POT1-ccdb (-BsaI) Gibson RP	GGGTACCGGGCC CCCCCtcgagctgcag accggtactagtGGTCT CTCTCAGAGGTGA GACGcgtcgacctgcaga ctgg	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT1 vector (by Gibson assembly)
POT1-ccdb (-BsaI) Gibson FP	GCTCCACCGCGGT GGCGGCCgcgaattcc ccgggtctagaGGTCTC TACCTGGCTAGAG ACGggatccggcttactaa aagccag	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT2 vector (by Gibson assembly)
POT2-ccdb (-BsaI) Gibson RP	GGGTACCGGGCC CCCCCtcgagctgcaga ccggtactagtGGTCTC TGCCTGAGGTGA GACGcgtcgacctgcaga ctgg	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT2 vector (by Gibson assembly)
POT3-ccdb (-BsaI)		

Gibson FP	GCTCCACCGCGGT GGCGGCCgcgaattc cccgggtctagaGGTCT CTAGGCGGCTAG AGACGggatccggctta ctaaaagccag	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT3 vector (by Gibson assembly)
POT1-ccdb (-BsaI)		
Gibson RP	GGGTACCGGGCC CCCCCtcgagctgcag accggtactagtGGTCT CTCTCAGAGGTGA GACGcgtcgacctgcaga ctgg	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT3 vector (by Gibson assembly)
POT3-ccdb (-BsaI)		
Gibson FP	GCTCCACCGCGGT GGCGGCCgcgaattc cccgggtctagaGGTCT CTAGGCGGCTAG AGACGggatccggctta ctaaaagccag	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT4 vector (by Gibson assembly)
POT4-ccdb (-BsaI)		
Gibson RP	GGGTACCGGGCC CCCCCtcgagctgcaga ccggtactagtGGTCTC TGGCAGAGGTGA GACGcgtcgacctgcaga ctgg	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT4 vector (by Gibson assembly)
POT5-ccdb(-BsaI)		
Gibson FP	GCTCCACCGCGGT GGCGGCCgcgaattc cccgggtctagaGGTCT CTTGCCGGCTAGA GACGggatccggcttact aaaagccag	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT5 vector (by Gibson assembly)
POT3-ccdb(-BsaI)		

Gibson RP	GGGTACCGGGCC CCCCCtcgagctgcag accggtactagtGGTCT CTCTCAGAGGTGA GACGcgtcgacctgcaga ctgg	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT5 vector (by Gibson assembly)
POT5-ccdb(-BsaI)		
Gibson FP	GCTCCACCGCGGT GGCGGCCgcgaattc cccgggtctagaGGTCT CTTGCCGGCTAGA GACGggatccggcttact aaaagccag	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT6 vector (by Gibson assembly)
POT6-ccdb (-BsaI)		
Gibson RP	GGGTACCGGGCC CCCCCtcgagctgcaga ccggtactagtGGTCTC TAGTGGAGGTGA GACGcgtcgacctgcaga ctgg	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT6 vector (by Gibson assembly)
POT7-ccdb (-BsaI)		
Gibson FP	GCTCCACCGCGGT GGCGGCCgcgaattcc ccgggtctagaGGTCTC TCACTGGCTAGAG ACGggatccggcttactaa aagccag	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT7 vector (by Gibson assembly)
POT1-ccdb (-BsaI)		
Gibson RP	GGGTACCGGGCC CCCCCtcgagctgcag accggtactagtGGTCT CTCTCAGAGGTGA GACGcgtcgacctgcaga ctgg	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT7 vector (by Gibson assembly)
POT7-ccdb (-BsaI)		

Gibson FP	GCTCCACCGCGGT GGCGGCCgcgaattcc ccgggtctagaGGTCTC TCACTGGCTAGAG ACGggatccggcttactaa aagccag	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT8 vector (by Gibson assembly)
POT8-ccdb (-BsaI) Gibson RP	GGGTACCGGGCC CCCCCtcgagctgcaga ccggtactagtGGTCTC TCGACGAGGTGA GACGcgtcgacctgcaga ctgg	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT8 vector (by Gibson assembly)
POT9-ccdb (-BsaI) Gibson FP	GCTCCACCGCGGT GGCGGCCgcgaattc cccgggtctagaGGTCT CTGTCGGGCTAG AGACGggatccggctta ctaaaagccag	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT9 vector (by Gibson assembly)
POT1-ccdb(-BsaI) Gibson RP	GGGTACCGGGCC CCCCCtcgagctgcag accggtactagtGGTCT CTCTCAGAGGTGA GACGcgtcgacctgcaga ctgg	Primers to amplify the ccdb cassette and clone the ccdb cassette into the POT9 vector (by Gibson assembly)
ccdb mut FP	atatggccagtgtgccggtat ccg	Primers to mutate the BsaI site within the ccdb cassette
ccdb mut RP	acttcttccccgataacggat accg	Primers to mutate the BsaI site within the ccdb cassette

HcKan_P ccdb FP-gib	GTGAGTTGATTgc agtccagttacgctggagtc CGTCTCgGGCTaG AGACCggatccggctta ctaaaagcc	Primers for replacing the RFP with the ccdB cassette in the HcKan_P parts acceptor vector
HcKan_P ccdb RP-gib	CTATggaggtcaggtatg atttaaatggtcagtCGT CTCgCATCaGAGA CCcgtcgacctgcagactg	Primers for replacing the RFP with the ccdB cassette in the HcKan_P parts acceptor vector
HcKan_O ccdb FP-gib	gtccagttacgctggagtccg tctcggatgagagaccggat ccggcttactaaaagccag	Primers for replacing the RFP with the ccdB cassette in the HcKan_O parts acceptor vector by Gibson assembly
HcKan_O ccdb RP-gib	caggtatgatttaaatggtca gtcgtctcggctaagagaccc gtcgacctgcagactgg	Primers for replacing the RFP with the ccdB cassette in the HcKan_O parts acceptor vector by Gibson assembly
HcKan_T ccdb FP - gib	GTGAGTTGATTgc agtccagttacgctggagtc CGTCTCgTAGCaG AGACCggatccggctta ctaaaagcc	Primers for replacing the RFP with the ccdB cassette in the HcKan_T parts acceptor vector by Gibson assembly
HcKan_T ccdb RP-gib	CTATggaggtcaggtatg atttaaatggtcagtCGT CTCgGAGGaGAGA CCcgtcgacctgcagactg	Primers for replacing the RFP with the ccdB cassette in the HcKan_T parts acceptor vector by Gibson assembly

YGO199	gcgtatatataccaatctaag tct	Primers for verifying the cloning of promoters into the Re2.8 reporter vector
YGO200	GTCAATTTACCGT AAGTAGCATC	Primers for verifying the cloning of promoters into the Re2.8 reporter vector
TRP flank FP	GACCAATCAGTAA AAATCAACGG	Primers for verifying the presence of the Z3EV transcription activator on the TRP locus
TRP flank RP	CTGTGAACATTCT CTTCAACAAG	Primers for verifying the presence of the Z3EV transcription activator on the TRP locus
McIsaac-HcK_FP	agcgtgGGTCTCgGG CTTTATattgaattttca aaaattcttacttt	Primers for the domestication of the pZ3EV promoter from McIsaac et al. (2013) to the YeastFab standard (into the HcKan_P vector) by Golden Gate assembly

McIsaac-HcK_RP	gtgctgGGTCTCaCATCTATAGttttttctccttgacgtta	Primers for the domestication of the pZ3EV promoter from McIsaac et al. (2013) to the YeastFab standard (into the HcKan_P vector) by Golden Gate assembly
RPS3_P FP	agcgtgGGTCTCgGGCTGGTCCACGTCAGTTCCACACAATAACATT	Primers for the domestication of pRPS3 to the YeastFab standard (into the HcKan_P vector)
RPS3_P RP	gtgctgGGTCTCaCATCTTTTGTAGTTTGTTTGCTGTTTTATTTTTCA	Primers for the domestication of pRPS3 to the YeastFab standard (into the HcKan_P vector)
RPS13 GG FP	agcgtgGGTCTCgGGCTGTATAGTCTATCTTAAACACTAAACTACC	Primers for the domestication of pRPS3 to the YeastFab standard (into the HcKan_P vector)
RPS13 GG RP	gtgctgGGTCTCaCATCTTTGACTGATTGTTGTTGATTGATTTAGTG	Primers for the domestication of pRPS3 to the YeastFab standard (into the HcKan_P vector)
HSP26_P GG FP	agcgtgGGTCTCgGGCTGCAGCAGCAACTCCGTGTGTACCCCTAAC	Primers for the domestication of pHSP26 to the YeastFab standard (into the HcKan_P vector)

HSP26_P GG RP	gtgctgGGTCTCaCATCGTTAATTTGTTTAGTTTGTTTGTTTGCTTT	Primers for the domestication of pHSP26 to the Yeast-Fab standard (into the HcKan_P vector)
pADH1-NeoHcKP Gib FP	GTGAGTTGATTgcagtccagttacgctggagtcCGTCTCgGGCTCctcgacgcatgcaacttc	Primers for the domestication of pADH1-neomycin riboswitch to the Yeast-Fab standard (into the HcKan_P vector) by Gibson assembly
pADH1-NeoHcKP Gib EP	CTATggaggtcaggtatgatttaaatggtcagtCGTCTCgCATCttcgacataggaccattaaaggacaag	Primers for the domestication of pADH1-neomycin riboswitch to the Yeast-Fab standard (into the HcKan_P vector) by Gibson assembly
vioA_0 HcKan_O GG FP	agcgtgGGTCTCaGATGGAAGCCAGAGATGATGC	Primers used for the domestication of vioA into the HcKan_O parts acceptor vector
vioA_0 HcKan_O GG RP	gtgctgGGTCTCgGTCCGGATCTGCATGCC	Primers used for the domestication of vioA into the HcKan_O parts acceptor vector
vioA_1 HcKan_O		

GG FP	agcgtgGGTCTCaGG ACTACTTAGTTGA GGCTATTGAAAG AGAAAG	Primers used for the domestication of vioA into the HcKan_O parts acceptor vector
vioA_1 HcKan_O GG RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCAGCCCTGCGG GTCC	Primers used for the domestication of vioA into the HcKan_O parts acceptor vector
vioA_0 HcKan_O Gibson FP	gtgagttgattgcagtccag ttacgctggagtccgtctcg GATGGAAGCCAG AGATGATG	Primers used for the domestication of vioA into the HcKan_O parts acceptor vector
vioA_1 HcKan_O Gibson RP	ggaggtcaggtatgattta aatggtcagtcgtctcggct aCTTATCGTCGTC ATCCTTGTAATCT CAGCCCTGCGGG TCCAAAG	Primers used for the domestication of vioA into the HcKan_O parts acceptor vector
vioB GG HcKan_O FP	agcgtgGGTCTCaGA TGGCAGCCTATCC GGTC	Primers used for the domestication of vioB into the HcKan_O parts acceptor vector
vioB GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCAACGACTTGCG TCTTCT	Primers used for the domestication of vioB into the HcKan_O parts acceptor vector
vioC GG HcKan_O FP	agcgtgGGTCTCaGA TGACTGAGTCACC GACTAC	Primers used for the domestication of vioC into the HcKan_O parts acceptor vector

vioC GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCATCGCTGCCCT AAAAC	Primers used for the domestication of vioC into the HcKan_O parts acceptor vector
vioD GG HcKan_O FP	agcgtgGGTCTCaGA TGACAGGTCCACT AGGTGC	Primers used for the domestication of vioD into the HcKan_O parts acceptor vector
vioD GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCATCCCGTGCCT CTTCGA	Primers used for the domestication of vioD into the HcKan_O parts acceptor vector
vioE GG HcKan_O FP	agcgtgGGTCTCaGA TGACCTCCACCCC ATGCGG	Primers used for the domestication of vioE into the HcKan_O parts acceptor vector
vioE GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCATCGTTCCGAT CCGCGT	Primers used for the domestication of vioE into the HcKan_O parts acceptor vector
vioF GG HcKan_O FP	agcgtgGGTCTCaGA TGACTGATTTGGA CTTCACTAG	Primers used for the domestication of viol into the HcKan_O parts acceptor vector
vioF GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC CTACTCCGTTGCG GCTAGT	Primers used for the domestication of viol into the HcKan_O parts acceptor vector

vioG GG HcKan_O FP	agcgtgGGTCTCaGA TGACAACCACGTC ACATGC	Primers used for the domestication of vioG into the HcKan_O parts acceptor vector
vioG GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCATACTGCTTCT CCTCCG	Primers used for the domestication of vioG into the HcKan_O parts acceptor vector
vioH GG HcKan_O FP	agcgtgGGTCTCaGA TGAGTGAACTAAC CGATACAAG	Primers used for the domestication of vioH into the HcKan_O parts acceptor vector
vioH GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCATCGCAGTTCT CCTCTC	Primers used for the domestication of vioH into the HcKan_O parts acceptor vector
vioI GG HcKan_O FP	agcgtgGGTCTCaGA TGACTGCTATTCC AGGCTC	Primers used for the domestication of vioI into the HcKan_O parts acceptor vector
vioI GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCATGACGCACCG CCAGTC	Primers used for the domestication of vioI into the HcKan_O parts acceptor vector
vioJ GG HcKan_O FP	agcgtgGGTCTCaGA TGAGACTGGACC GACACC	Primers used for the domestication of vioJ into the HcKan_O parts acceptor vector

vioJ GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCAGACGTCGGTC CTTGG	Primers used for the domestication of vioJ into the HcKan_O parts acceptor vector
vioK GG HcKan_O FP	agcgtgGGTCTCaGA TGACCATCCAAGC AGTTCC	Primers used for the domestication of vioK into the HcKan_O parts acceptor vector
vioK GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCACCGGTCTGCT CCGAG	Primers used for the domestication of vioK into the HcKan_O parts acceptor vector
vioL GG HcKan_O FP	agcgtgGGTCTCaGA TGACCGAACCCAG AGGTCTG	Primers used for the domestication of vioL into the HcKan_O parts acceptor vector
vioL GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCACGTCGTGACC ACCC	Primers used for the domestication of vioL into the HcKan_O parts acceptor vector
vioM GG HcKan_O FP	agcgtgGGTCTCaGA TGAGAACGCGTAT TGCTTTAAG	Primers used for the domestication of vioM into the HcKan_O parts acceptor vector
vioM GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCATCGTGTGGG CTCCTCG	Primers used for the domestication of vioM into the HcKan_O parts acceptor vector

vioN GG HcKan_O FP	agcgtgGGTCTCaGA TGAACGACACCCC TGCG	Primers used for the domestication of vioN into the HcKan_O parts acceptor vector
vioN GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCATGCGCGGGC GCTCAAG	Primers used for the domestication of vioN into the HcKan_O parts acceptor vector
vioO GG HcKan_O FP	agcgtgGGTCTCaGA TGCCAACCACAGG TACAGC	Primers used for the domestication of vioO into the HcKan_O parts acceptor vector
vioO GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCACTGGGGCTCT CCTCG	Primers used for the domestication of vioO into the HcKan_O parts acceptor vector
vioP GG HcKan_O FP	agcgtgGGTCTCaGA TGGAGCCCGACT GGAGAAG	Primers used for the domestication of vioP into the HcKan_O parts acceptor vector
vioP GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCAGGCGCATGCC TGACCC	Primers used for the domestication of vioP into the HcKan_O parts acceptor vector
vioQ GG HcKan_O FP	agcgtgGGTCTCaGA TGCCGACGTGTA GA	Primers used for the domestication of vioQ into the HcKan_O parts acceptor vector

vioQ GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCATCTAGATTCC TTGAAA	Primers used for the domestication of vioQ into the HcKan_O parts acceptor vector
vioS GG HcKan_O FP	agcgtgGGTCTCaGA TGACTATGAAATG CAGAG	Primers used for the domestication of vioS into the HcKan_O parts acceptor vector
vioS GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TCATCTCAAGTTT AAACCT	Primers used for the domestication of vioS into the HcKan_O parts acceptor vector
vph GG HcKan_O FP	agcgtgGGTCTCaGA TGGAGACTTCGCC CACC	Primers used for the domestication of vph into the HcKan_O parts acceptor vector
vph GG HcKan_O RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC CTACCGGTAGCC GCTTAAT	Primers used for the domestication of vph into the HcKan_O parts acceptor vector
Transferase GG FP	agcgtgGGTCTCaGA TGAAAATATACG GAATATACATGG ATAG	Primers used for the domestication of pptase into the HcKan_O parts acceptor vector
Transferase GG RP	gtgctgGGTCTCgGC TACTTATCGTCGT CATCCTTGTAATC TTATAAAAGCTCC TCGTACGATAC	Primers used for the domestication of pptase into the HcKan_O parts acceptor vector

vioA scr 0R	CTAGGAGCAGTC CACCATG	Sequencing primers for vioA, for primer walking
vioA scr 1F	GTTCATGACCGAT TACCTTGTG	Sequencing primers for vioA, for primer walking
vioA scr 1R	GCGCCTGCTAAAT GTAGTG	Sequencing primers for vioA, for primer walking
vioA scr 2F	CTGGCCCAAATCT CTGAAAC	Sequencing primers for vioA, for primer walking
vioA scr 2R	CTCTCACTGATAA ACCAGGTAAG	Sequencing primers for vioA, for primer walking
vioA scr 3F	CCGTCTTCCAGGT AATAGCTC	Sequencing primers for vioA, for primer walking
vioA scr 3R	GAACAGGCAGGG TTTCACC	Sequencing primers for vioA, for primer walking
vioA scr 4F	CTCTCCTTCGACG TGTCC	Sequencing primers for vioA, for primer walking
vioA scr 4R	GCCATTAGATCGT GCAACG	Sequencing primers for vioA, for primer walking
vioA scr 5F	CAATTATGGCCAG GTACCGC	Sequencing primers for vioA, for primer walking
vioF chc 0R	CGGTAACTTCATA CGCAGT	Sequencing primers for vioF, for primer walking

vioF chc 1F	GTTTAACATCTAC GGGCCTAC	Sequencing primers for vioF, for primer walking
vioF chc 1R	GTTCTTCCGTTCT AGCTCTAAATC	Sequencing primers for vioF, for primer walking
vioF chc 2F	GCGACATGAGGT TTTAAGAAC	Sequencing primers for vioF, for primer walking
vioG chc 0R	CTGTCCTTCTGTA ACACCTG	Sequencing primers for vioG, for primer walking
vioG chc 1F	CTCAACACCATCT ACGGG	Sequencing primers for vioG, for primer walking
vioG chc 1R	CAAACAATAATG GACGCACTC	Sequencing primers for vioG, for primer walking
vioG chc 2F	GATTTGACTTACC GGTCCTG	Sequencing primers for vioG, for primer walking
vioA 200bp FP	TGGTTACTGGCCC AAATCTC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioA in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioA 200bp RP	CTTAATGCTTCGG CAGGAAC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioA in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioA_0 200bp FP	TGGTTACTGGCCC AAATCTC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioA in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioA_0 200bp RP	CTTAATGCTTCGG CAGGAAC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioA in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioA_1 200bp FP	CTATGCTGTACGT GCCGAAG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioA in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioA_1 200bp RP	ATCGGAACCTGAA TCACCAG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioA in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioB 200bp FP	AGACAGGCTGTT GCATGTTG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioB in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioB 200bp RP	TAAGCCCATTCTG TCCAACC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioB in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioC 200bp FP	AGAGCAGCACCCC ATCTTAG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioC in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioC 200bp RP	GTCGTAGCGAGG TTGGAAAG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioC in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioD 200bp FP	CCCATCAGGTGCA TGTGTAG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioD in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioD 200bp RP	ACCCTTAGACCTG GCAAACC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioD in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioE 200bp FP	ACTTACCGGGTT GGGGTATC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioE in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioE 200bp RP	CCACCGGTCCAAT AACAAAC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioE in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioF 200bp FP	TTCCCACTCGGAA TCGTATC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioF in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioF 200bp RP	ATCCTGCCAAAAG TTCATCG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioF in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioG 200bp FP	GCTGAGGTTTTA GGGGTTCC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioG in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioG 200bp RP	ATCAGCAGCGGCT AATTGAC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioG in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioH 200bp FP	CAAGATCACCGA GGGAATTG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioH in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioH 200bp RP	CCCGTTAAAACGA CAAATGG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioH in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioI 200bp FP	GCATGACGATGG AGATGATG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioI in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioI 200bp RP	TTCATCTTCTGGG CGAACTC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioI in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioJ 200bp FP	GGGCAAGGGAAT TAGGTAGG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioJ in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioJ 200bp RP	ACGGCTGCTAAAT CTGAACC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioJ in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioK 200bp FP	ATGACCATCCAAG CAGTTCC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioK in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioK 200bp RP	GTAAGCACCATG GGAGTTGG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioK in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioL 200bp FP	AACACAAGGATT GGCTGACC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioL in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioL 200bp RP	GTTTCTGCAAGCA CAGCATC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioL in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioM 200bp FP	AGATGGTCATGT CCCCTTTG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioM in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioM 200bp RP	CACCTTCTCTTCC CAGTTCG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioM in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioN 200bp FP	CGCATACCAGGTC GTTTTG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioN in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioN 200bp RP	TGATCTAAACACG CCTGACG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioN in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioO 200bp FP	GCTCCAGCTCTGA TCACTCC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioO in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioO 200bp RP	TGGAACATAAGCT GCACCAG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioO in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioP 200bp FP	ATTAGATGCCGG AGTTCGTG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioP in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioP 200bp RP	GTCCTGAGCTTCG GCTAATG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioP in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioQ 200bp FP	TTTGTGGGAAACT GTCGATG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioQ in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioQ 200bp RP	GCAAAGGTGCCT CGTATAGC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioQ in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vioS 200bp FP	GTAGCCTACGGC CTTGTCTG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioS in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vioS 200bp RP	ACATGGCCCATAG AGTCAGC	Primers used for PCR, RT-PCR and qPCR verification for the presence of vioS in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
vph 200bp FP	GCAAAGAGCAAG ACAACACG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vph in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

vph 200bp RP	ACTCGGGATACC GTCAGATG	Primers used for PCR, RT-PCR and qPCR verification for the presence of vph in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
ppt 200bp FP	GAAAAACGGGAG AAATGCAG	Primers used for PCR, RT-PCR and qPCR verification for the presence of ppt in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)

ppt 200bp RP	ATCTGCCGGAGT GAGAAATG	Primers used for PCR, RT-PCR and qPCR verification for the presence of ppt in the YAC. This primer amplifies a 200 bp region of the gene. The primer was designed by Dr. Almer van der Sloot from the Tyers laboratory (University of Montreal)
nPAL1 GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GGAAATTAACGG TGCCCAC	Primers used for the domestication of atPAL1 into the HcKan_O parts acceptor vector
nPAL1 GG RP	gtgctgGGTCTCgGC TAACATATTGGAA TGGGAGCACC	Primers used for the domestication of atPAL1 into the HcKan_O parts acceptor vector
n4CL3 GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GATCACTGCAGCT CTACACG	Primers used for the domestication of at1-4CL into the HcKan_O parts acceptor vector
n4CL3 GG RP	gtgctgGGTCTCgGC TATATCATGATGT AGATCTTTTTCAA GTGCAG	Primers used for the domestication of at1-4CL into the HcKan_O parts acceptor vector

nChsyn GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GGTGACTGTAGA AGAAGTCAGGAA AG	Primers used for the domestication of haCHS into the HcKan_O parts acceptor vector
nChsyn GG RP	gtgctgGGTCTCgGC TAGTTAATGGCGA CACTGTGAAG	Primers used for the domestication of haCHS into the HcKan_O parts acceptor vector
nTal1 GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GACCCTGCAGTCT CAAACG	Primers used for the domestication of rcTAL into the HcKan_O parts acceptor vector
nTal1 GG RP	gtgctgGGTCTCgGC TATGCCGGAGGA TCAGCG	Primers used for the domestication of rcTAL into the HcKan_O parts acceptor vector
nC4H GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GGACCTTTTATTG CTGGAGAAGTCTT TAATC	Primers used for the domestication of atC4H into the HcKan_O parts acceptor vector
nC4H GG RP	gtgctgGGTCTCgGC TAACAGTTCCTTG GTTTCATAACG	Primers used for the domestication of atC4H into the HcKan_O parts acceptor vector

nCPR1 GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GCCGTTTGGAATA GACAACACAG	Primers used for the domestication of scCPR1 into the HcKan_O parts acceptor vector
nCPR1 GG RP	gtgctgGGTCTCgGC TACCAGACATCTT CTTGGTATCTACC	Primers used for the domestication of scCPR1 into the HcKan_O parts acceptor vector
nCHI-1 GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GAATCCATCGCCG TCCG	Primers used for the domestication of csCHI into the HcKan_O parts acceptor vector
nCHI_1 GG RP	gtgctgGGTCTCgGC TATTTCATCTTAT CACTAGTTACATT CAACAGTG	Primers used for the domestication of csCHI into the HcKan_O parts acceptor vector
nCHI-2 GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GGCAGCTGCAGC GGCTG	Primers used for the domestication of plCHI into the HcKan_O parts acceptor vector
nCHI-2 GG RP	gtgctgGGTCTCgGC TAGACTATAATGC CGTGGCTCAATAC AG	Primers used for the domestication of plCHI into the HcKan_O parts acceptor vector
nCHI-3 GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GGCGGCTGTTTCT GAAGTAGAAG	Primers used for the domestication of osCHI into the HcKan_O parts acceptor vector

nCHI-3 GG RP	gtgctgGGTCTCgGC TACGCGGAGACA GGTGC	Primers used for the domestication of osCHI into the HcKan_O parts acceptor vector
nCHI-4 GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GTGCTGCAGCAT CCTCC	Primers used for the domestication of psCHI into the HcKan_O parts acceptor vector
nCHI-4 GG RP	gtgctgGGTCTCgGC TAATTCCCAATCT TGAAAGTACCCTC	Primers used for the domestication of psCHI into the HcKan_O parts acceptor vector
nCHI-5 GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GTCTCCTCCAGTG TCCGTTAC	Primers used for the domestication of ph1CHI into the HcKan_O parts acceptor vector
nCHI-5 GG RP	gtgctgGGTCTCgGC TAGACTCCAATCA CTGGAATAGTAG ATTTC	Primers used for the domestication of ph1CHI into the HcKan_O parts acceptor vector
nCHI-6 GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATA AGAGCCCATCAGT GTCTGTTAC	Primers used for the domestication of ph2CHI into the HcKan_O parts acceptor vector
nCHI-6 GG RP	gtgctgGGTCTCgGC TAAAATAAACTAT TCTTGAAAAGTTC AGACATTCTTG	Primers used for the domestication of ph2CHI into the HcKan_O parts acceptor vector

nCPR-AT GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GACTTCTGCTTTG TATGCTTCC	Primers used for the domestication of atCPR1 into the HcKan_O parts acceptor vector
nCPR-AT GG RP	gtgctgGGTCTCgGC TACCAGACATCTC TGAGGTATCTTC	Primers used for the domestication of atCPR1 into the HcKan_O parts acceptor vector
nCHI-AT GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GGGCACTGAAAT GGTCATGG	Primers used for the domestication of atCHIinto the HcKan_O parts acceptor vector
nCHI-AT GG RP	gtgctgGGTCTCgGC TAGGTTAAAACTG CGGAGATTGAAT C	Primers used for the domestication of atCHIinto the HcKan_O parts acceptor vector
nCHS-AT GG FP	agcgtgGGTCTCaGA TGGATTACAAGG ATGACGACGATAA GGTAATGGCTGG TGCGTCTTC	Primers used for the domestication of atCHS into the HcKan_O parts acceptor vector
nCHS-AT GG RP	gtgctgGGTCTCgGC TAGAGGGGAACG CTGTGC	Primers used for the domestication of atCHS into the HcKan_O parts acceptor vector

seq0 hom F	GCGATCTGCGATA TATTTATG	Primers to amplify the 2-5 kb chunks of the naringenin pathway. These primers bind to the homology arms of YFass 1, therefore they will amplify the multi-transcription units assembled into YFass 1. This segment will be used as a template for sequencing by primer walking
seq1 hom R	GTCCTCGTAAGTA TTATGTAAGG	Primers to amplify the 2-5 kb chunks of the naringenin pathway. These primers bind to the homology arms of YFass 1, therefore they will amplify the multi-transcription units assembled into YFass 1. This segment will be used as a template for sequencing by primer walking

seq1 hom F	GAATACTATGATT CCTGTGGAC	Primers to amplify the 2 to 5 kb chunks of the naringenin pathway. These primers bind to the homology arms of YFass 2, therefore they will amplify the multi-transcription units assembled into YFass 2. This segment will be used as a template for sequencing by primer walking
seq2 hom R	CTCGAGAGTCATT GTTAAAGG	Primers to amplify the 2 to 5 kb chunks of the naringenin pathway. These primers bind to the homology arms of YFass 2, therefore they will amplify the multi-transcription units assembled into YFass 2. This segment will be used as a template for sequencing by primer walking

seq2 hom F	CATGATGTAAAG ATATGTCAGTAAC	Primers to amplify the 2-5 kb chunks of the naringenin pathway. These primers bind to the homology arms of YFass 3, therefore they will amplify the multi-transcription units assembled into YFass 3. This segment will be used as a template for sequencing by primer walking
seq3 hom R	CTAGGACAACAG ATGTTCATC	Primers to amplify the 2-5 kb chunks of the naringenin pathway. These primers bind to the homology arms of YFass 3, therefore they will amplify the multi-transcription units assembled into YFass 3. This segment will be used as a template for sequencing by primer walking

seq3 hom F	GTAAGTGGGTCA TTTTCTATAACTC	Primers to amplify the 2-5 kb chunks of the naringenin pathway. These primers bind to the homology arms of YFass 4, therefore they will amplify the multi-transcription units assembled into YFass 4. This segment will be used as a template for sequencing by primer walking
seq4 hom R	CTTTTGTGGTGGT AAATGG	Primers to amplify the 2-5 kb chunks of the naringenin pathway. These primers bind to the homology arms of YFass 4, therefore they will amplify the multi-transcription units assembled into YFass 4. This segment will be used as a template for sequencing by primer walking

seq4 hom F	GATTACAGAACTT ATATGGTTCAC	Primers to amplify the 2-5 kb chunks of the naringenin pathway. These primers bind to the homology arms of YFass 5, therefore they will amplify the multi-transcription units assembled into YFass 5. This segment will be used as a template for sequencing by primer walking
seq25 hom R	CTTTTACGCTTGG ACTGTC	Primers to amplify the 2-5 kb chunks of the naringenin pathway. These primers bind to the homology arms of YFass 5, therefore they will amplify the multi-transcription units assembled into YFass 5. This segment will be used as a template for sequencing by primer walking
4CL3_0R scr	GAGGTAAAGATT TGGGTTATCTCC	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing

4CL3_1F scr	GTGTAGTTTTGACCCATAAGAGC	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
4CL3_PRO_RP	GAAAATGCGAGGCGTAGGAG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
4CL3_TER FP	GGAGGTGATAAAGTTTAAAGGTTTCC	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
C4H_0R scr	CTACCTGTAGGCTTAGAACTCG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
C4H_1F scr	GAAGGATTGCGCTTTTCAAGAAG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
C4H_PRO RP	CCTGGAGGTAGCTTCAATTTCTTG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
C4H_TER FP	CCTCCAGGACAGTCTAAAGTG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing

CHI-6_PRO_RP	CGCCACCAAGGAA CAAAG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
CHI-6_TER_FP	GAAGCAGTGCTA GAAACGATTATT G	Extra sequencing primers to complete the coverage of JAy081 pathway sequencing
CHS_PRO_RP	GCTTGGTCAACG CAGTTC	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
CHS_TER_FP	CTTAAGACCACAG GGGAAGG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
CPR_PRO_RP	CGGACATTAATAA TTCCTTGATGGAG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
CPR_TER FP	GTTGGTATATTGT CCCGTGG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
CPR1_0R scr	CAGGGGCTTCAA GTCAAAAATG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing

CPR1_1F scr	CCCATTAGAAAAG GTCGAACAATTC	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
PAL_0R scr	CAGCCAAAACAGA TAACACATTCG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
PAL_1F scr	CAATGGTACCGCA GTGG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
PAL_1R scr	GTAATTCACCATT TACCCCAGTAG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
PAL_2F scr	GAGAATTTGAGA CAGACTGTGAAG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
PAL1_PRO RP	GTCCTCCGCGTTG ATCAC	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
PAL1_TER FP	CAGGAGAAGAGT TCGACAAGG	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing

TAL_0R scr	CATCAAGAGCTTG TGCCAAC	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
TAL_1F scr	GCGTTTGCTGAG TTAAGACC	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
TAL_PRO RP	GGGGTGACAGAT ATTCTTGACC	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing
TAL_TER FP	GGTAAGTCCCGG TTTAAGAGC	Additional sequencing primers to complete the coverage of JAy081 pathway sequenciing

Appendix B

Strains used in this study

Table B.1 E coli competent cells used in this study

Name of strain	Genotype	Comment	Manufacturer and cat no.
ccdB Survival 2	F-mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 araΔ139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG fhuA::IS2	Used for propagating plasmids containing the ccdB toxic gene	Invitrogen cat no. A10460
DH5α	F– ϕ80lacZΔ M15 Δ (lacZYA-argF) U169 recA1 endA1 hsdR17 (rK– mK+) phoA supE44 λ- thi–1 gyrA96 relA1	Used for the cloning and propagation of plasmids in *E. coli*cells	Thermo scientific cat no. EC0112
DH10B	F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZ ΔM15 ΔlacX74 recA1 endA1 araD139 Δ (ara-leu)7697 galU galK λ– rpsL(StrR) nupG	Used for the cloning and propagation of plasmids in *E. coli*cells	Thermo scientific cat no. EC0113
Mach1 T1R	F– φ80lacZΔM15 ΔlacX74 hsdR(rK-, mK+) ΔrecA1398 endA1 tonA	Used for the cloning and propagation of plasmids in *E. coli*cells. Shorter doubling time.	Invitrogen cat no. C862003

Table B.2 Table of bacterial used in this study

Name	Information	Manufacturer and catalogue number or origin of strain	Selection media
Bacillus subtilis 168	*Bacillus subtilis* (Ehrenberg) Cohn strain 168 (ATCC 23857)	Sent by the Wright group (McMaster university, Canada)	None
Micrococcus luteus	*Micrococcus luteus* (ATCC 4698)	Obtained from the University of Edinburgh teaching lab	None
Mycobacterium smegmatis	*Mycobacterium smegmatis* (ATCC 607)	Purchased from Public Health England Culture collections (NCTC 7017)	None

Table B.3 Table of yeast strains used in this study

Strain ID	Name	Information	Genotype	Selection media
SASy31	SASy31	Background strain constructed by a collaborator, Susan Moore	MATα, ΔHIS3, Δ LEU2, Δ TRP1, Δ URA3, Lys. Z3EV transcription factor integrated into the Trp locus.	
SASy35	SASy35	Background strain constructed by a collaborator, Susan Moore	MATa, Δ HIS3, ΔLEU2, Δ TRP1, Δ URA3, Lys. Z3EV transcription factor integrated into the Trp locus.	
BY4741	BY4741	Wild type yeast strain used in this study. (Winston et al. 1995)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	
JAy109	pADH1-neo pWHE601	BY4741 containing the pADH1-neo in the pWHE601 plasmid. Gift from Prof. Beatrix Suess from Technical University of Darmstadt, Germany.	Same as BY4741	URA
JAy508	pADH1-neo Re2.8	BY4741 containing the pADH1-neo in the Re2.8 plasmid.	Same as BY4741	URA, LEU
JAy509	McIsaac Re2.8	SASy31 transformed with the pZ3EV-Re2.8 reporter vector.	Same as SASy31	URA, LEU
JAy008	Re2.8	BY4741 containing the Re2.8 dual fluorescence reporter vector previously constructed by a collaborator - Dai lab	Same as BY4741	URA, LEU

JAy001	Vio1	SASy31 strain containing all 20 Viomycin biosynthetic enzyme genes on a linear YAC (YAC10)	Same as SASy31	HIS, LEU
JAy021	Minvio6	SASy31 strain containing minimal viomycin biosynthetic enzyme genes in a circular YAC (YAC10)	Same as SASy31	HIS, LEU, URA
JAy096	pYAC10	SASy31 strain containing the circular YAC (YAC10)	Same as SASy31	HIS, LEU, URA
JAy078	NC2	SASy31 strain containing the full naringenin biosynthetic enzyme genes in a circular YAC (YAC10) with CHI-2	Same as SASy31	HIS, LEU, URA
JAy079	NC3	SASy31 strain containing the full naringenin biosynthetic enzyme genes in a circular YAC (YAC10) with CHI-3	Same as SASy31	HIS, LEU, URA
JAy080	NC4	SASy31 strain containing the full naringenin biosynthetic enzyme genes in a circular YAC (YAC10) with CHI-4	Same as SASy31	HIS, LEU, URA
JAy081	NC6	SASy31 strain containing the full naringenin biosynthetic enzyme genes in a circular YAC (YAC10) with CHI-6	Same as SASy31	HIS, LEU, URA
JAy092	Nar-AT	SASy31 strain containing the full naringenin biosynthetic enzyme genes in a circular YAC (YAC10) with all A. thaliana genes except for TAL	Same as SASy31	HIS, LEU, URA

JAy093	Truncated naringenin	SASy31 strain containing the truncated naringenin pathway	Same as SASy31	HIS, LEU, URA

Appendix C

Supplementary figures and tables - for Chapter 3

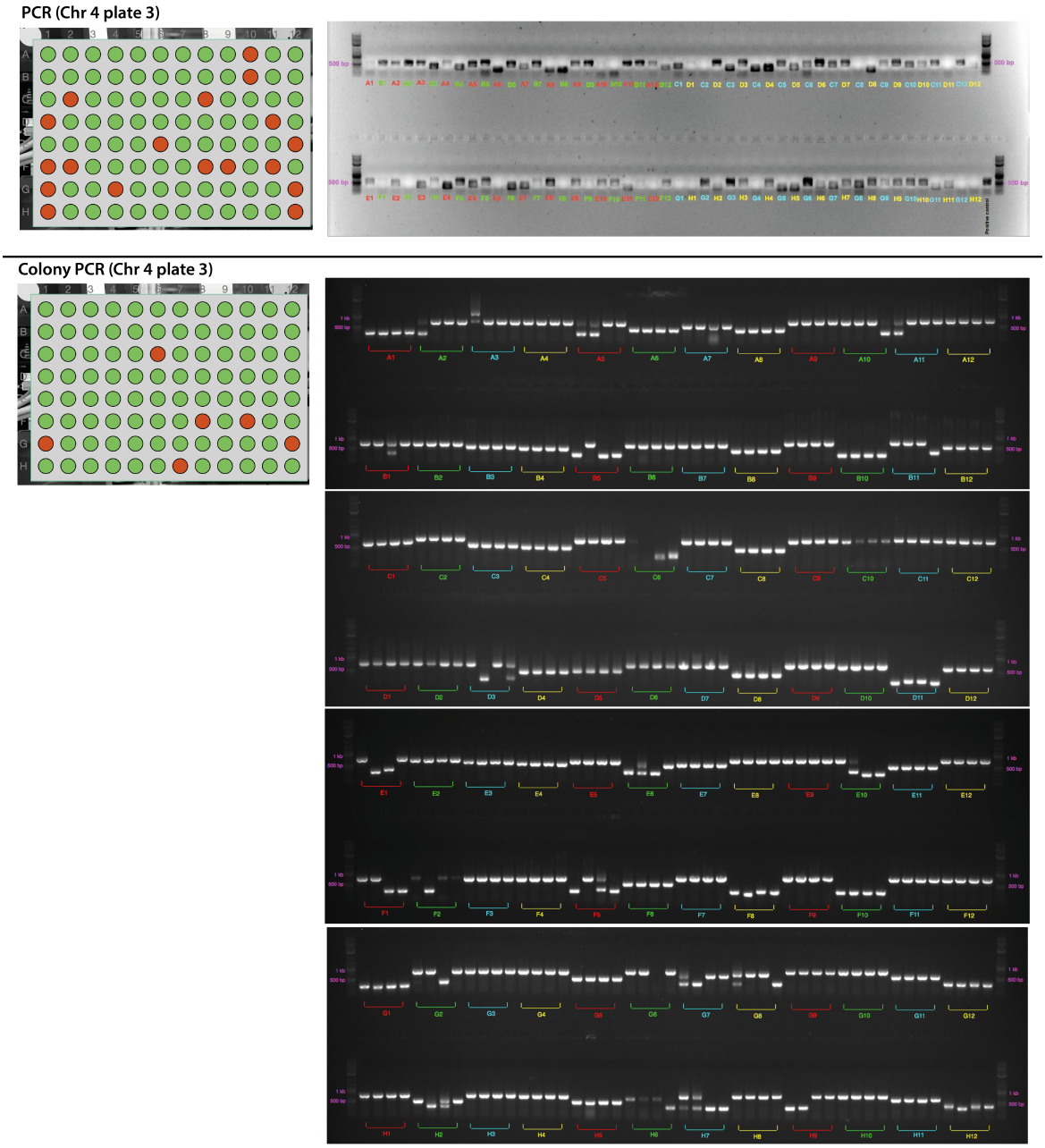


Figure C.1

Scoring of promoters in the PCR and colony PCR step using the PCR gels and colony PCR agarose gels of promoters on Chr 4 plate 3

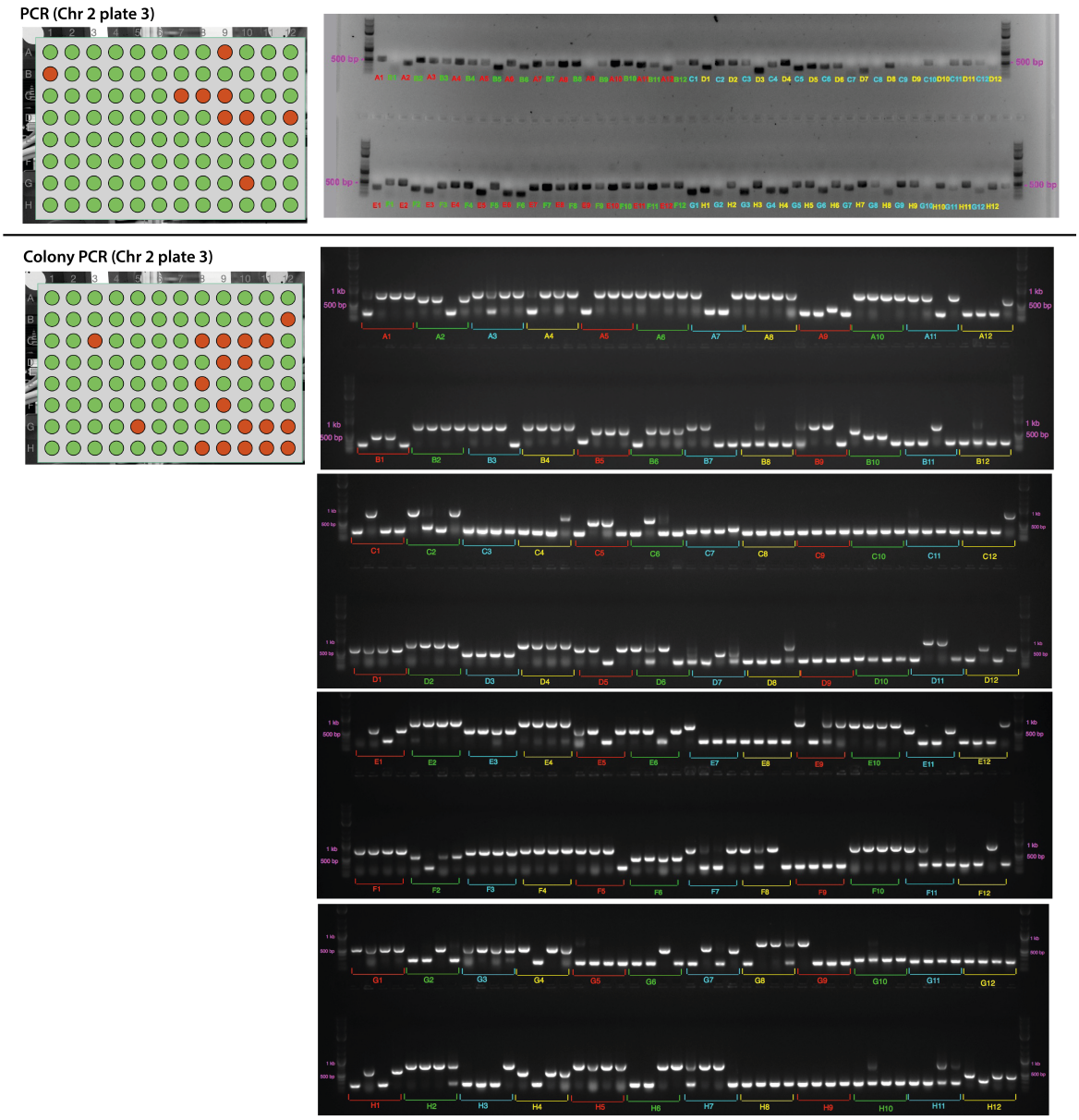


Figure C.2

Scoring of promoters in the PCR and colony PCR step using the PCR gels and colony PCR agarose gels of promoters on Chr 2 plate 3

Appendix D

Supplementary figures and tables - for Chapter 4

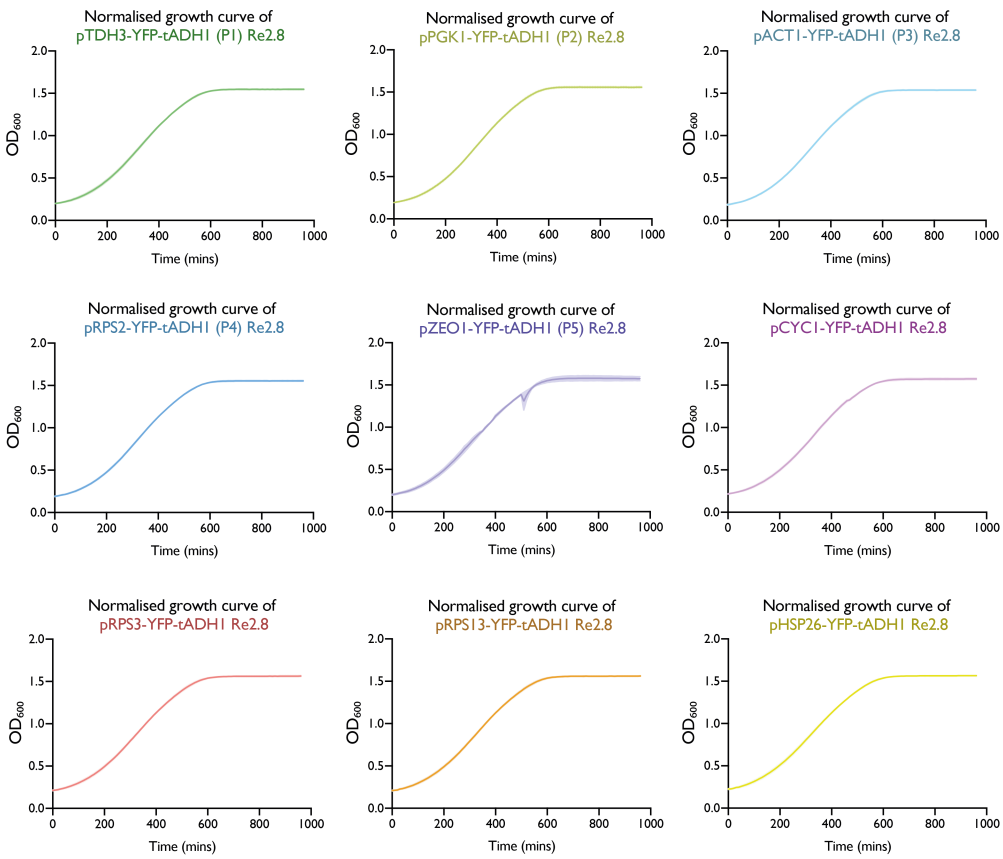


Figure D.1

Normalised growth curves of the 9 promoters characterised using the plate reader over 20 hours

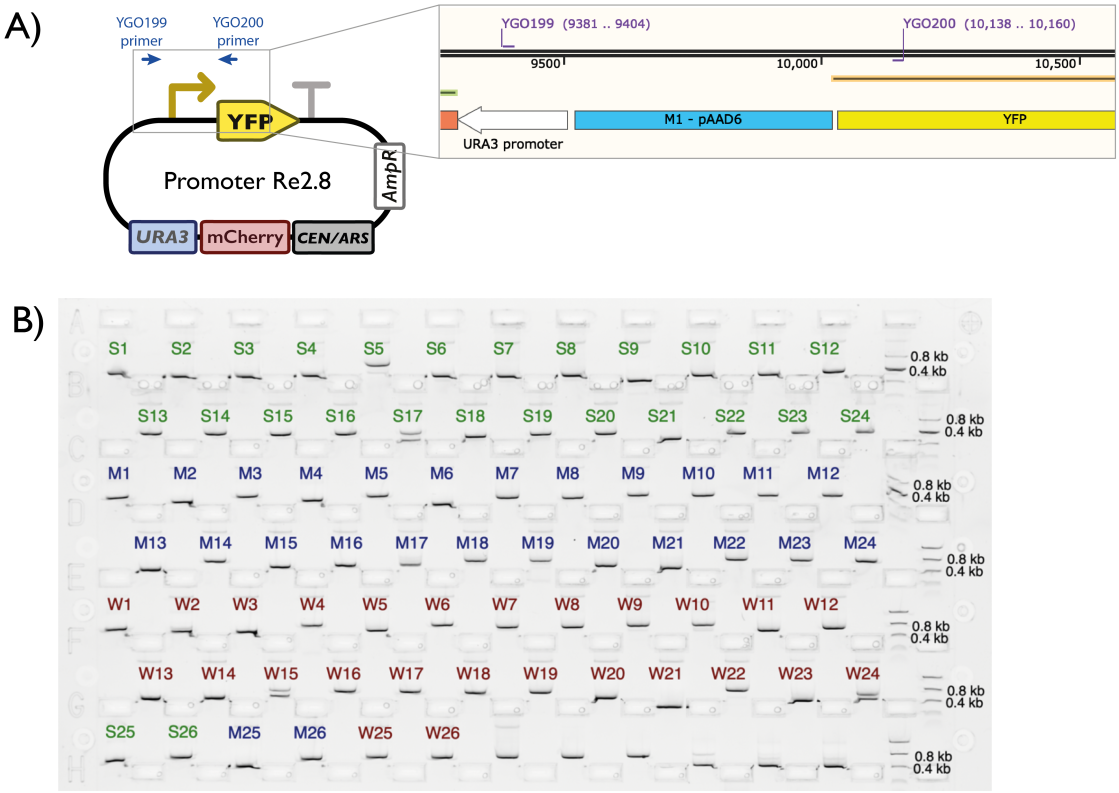


Figure D.2

A) Location of the primer binding sites of YGO199 and YGO200 for the PCR verification of promoter-Re2.8 assemblies. B) PCR verification results, visualised on an E-gel (Invitrogen), for the cloning of 78 promoters into the Re2.8 dual fluorescence reporter vector.

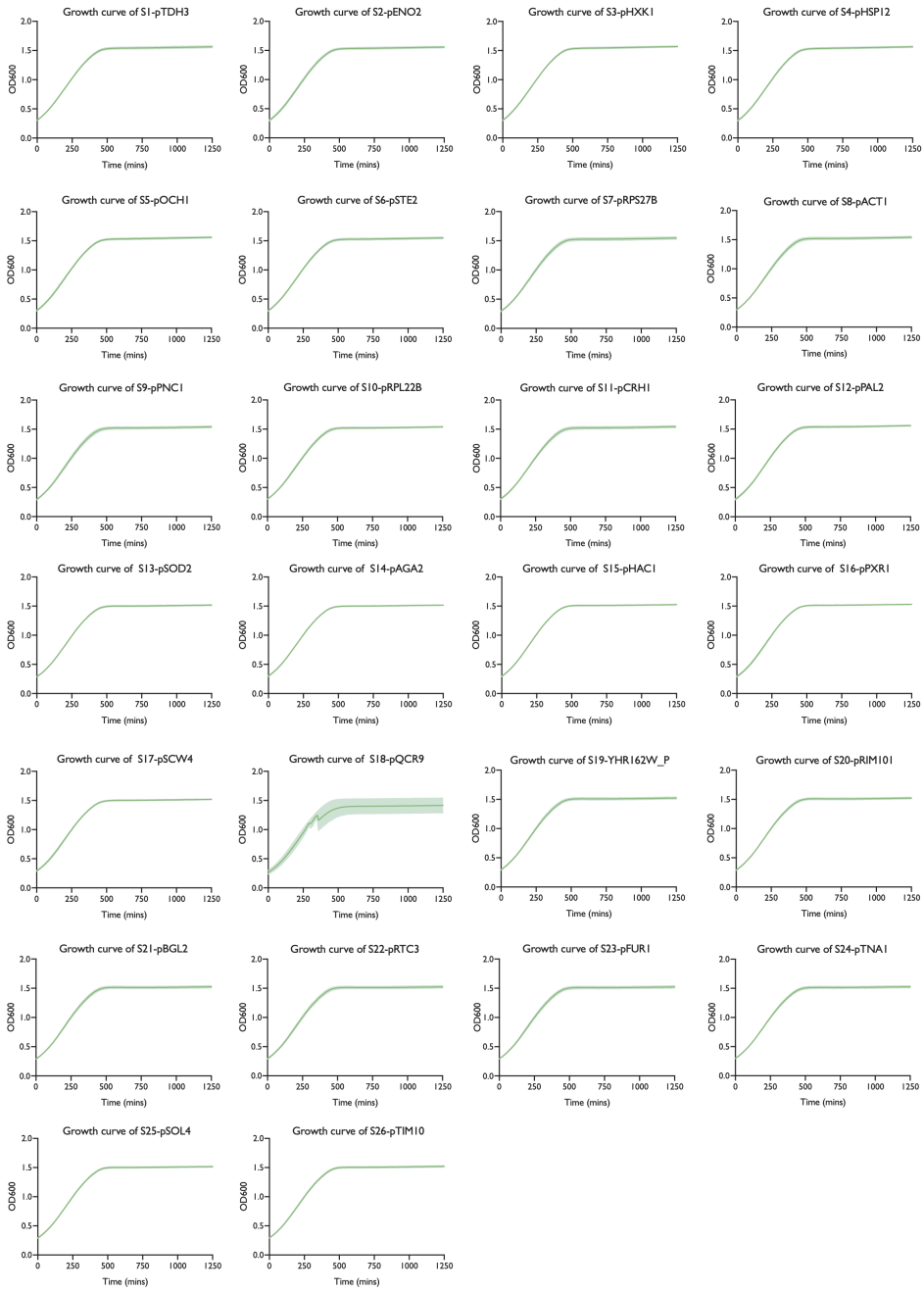


Figure D.3

Normalised growth curves of all 25 'strong' promoters across three experimental replicates (n=3). Growth curves were normalised by the peak growth rate to account for the difference in lag time.

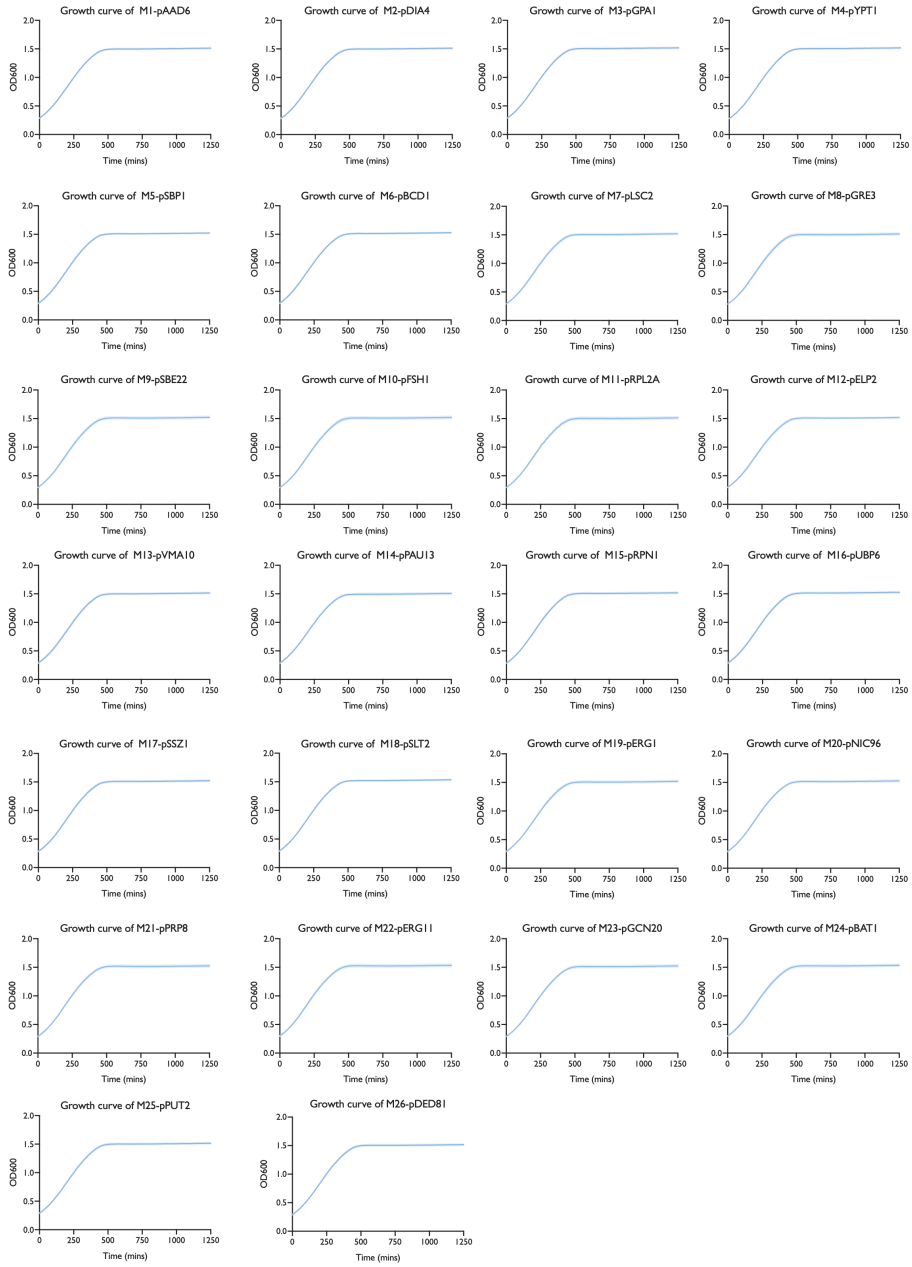


Figure D.4

Normalised growth curves of all 26 ‘moderate strength’ promoters across three experimental replicates (n=3). Growth curves were normalised by the peak growth rate to account for the difference in lag time.

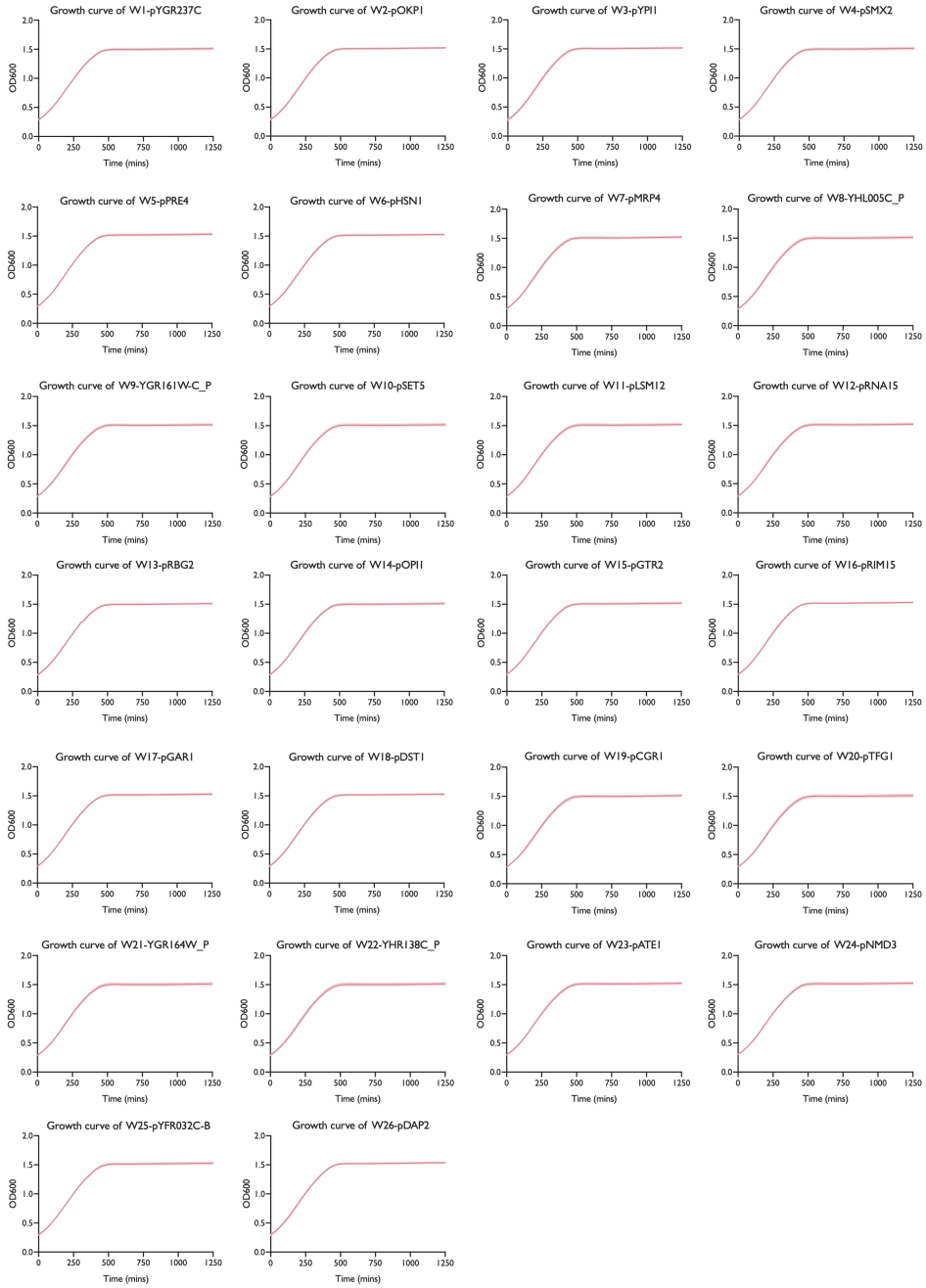


Figure D.5

Normalised growth curves of all 23 'weak' promoters across three experimental replicates (n=3). Growth curves were normalised by the peak growth rate to account for the difference in lag time.

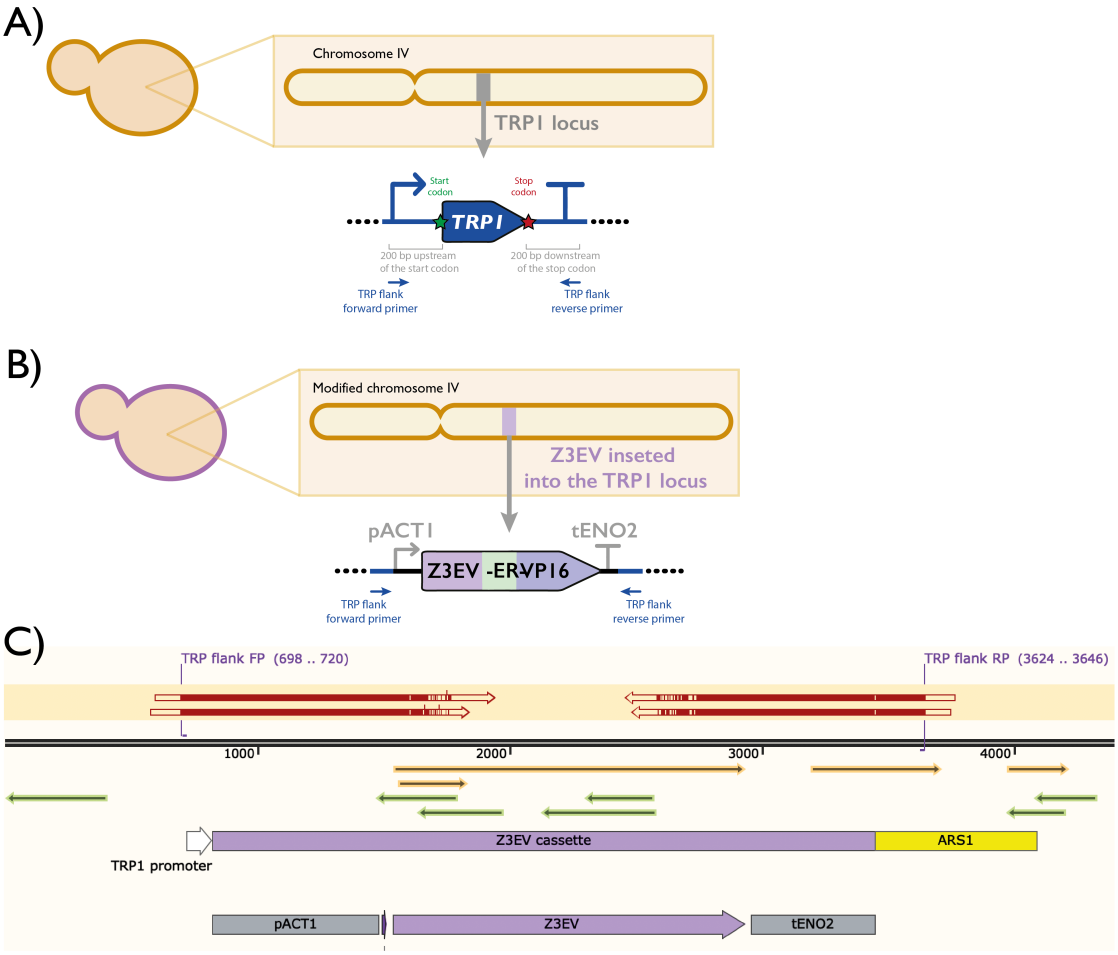


Figure D.6

Verification of the knock-in of the Z3EV transcription activator into the TRP1 locus. The introduction of Z3EV was carried out by Susan Moore from the Tyers laboratory at the University of Montreal. A) The location of the verification primers for the TRP1 locus. The forward primer was designed 200 bp upstream of the TRP1 start codon and the reverse primer was designed 200 bp downstream of the TRP1 stop codon (Table of primers). B) The primers were used to amplify the Z3EV gene knocked into the TRP1 locus in the SASY31 yeast strain and subcloned into the TOPO blunt II vector. This plasmid was sent for sequencing verification using M13 FP and RP primers (Table of primers). C) Sequencing result of two clones, showing the presence of the Z3EV insert in the correct orientation in the TRP1 locus.

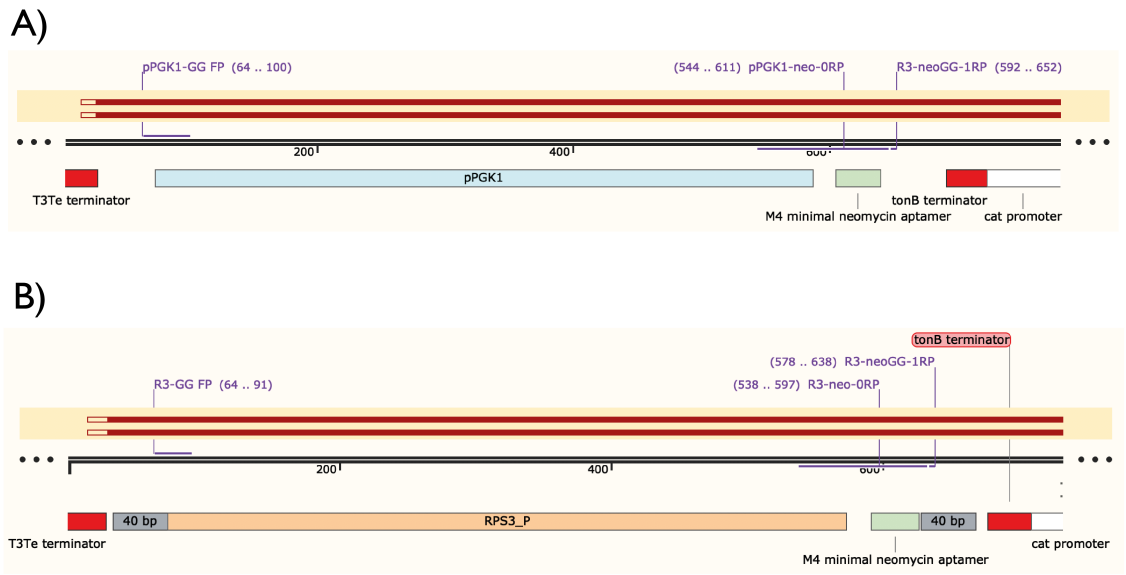


Figure D.7

The construction of modified variants of the neomycin riboswitch. A) The M4 neomycin aptamer was inserted into the pPGK1 5'UTR and cloned into the YeastFab HcKan_P vector by nested PCR and Gibson assembly. Sequencing verification shows the successful assembly of this riboswitch. B) The M4 neomycin aptamer was inserted into the pRPS3 5'UTR and cloned into the YeastFab HcKan_P vector by nested PCR and Gibson assembly. Sequencing verification shows the successful assembly of this riboswitch.

Table D.1 26 'strong' promoters and the characterised activity from the Guo et al. (2015) publication. The corresponding ORFs and key functions as described in the Saccaromyces genome database (SGD - https://www.yeastgenome.org) of each of the 26 promoters in this library.

Promoter name	Promoter ID	ORF name and aliases	Key functions of ORF (From SGD)	Characterised promoter activity (Guo et al., 2015)
S1_TDH3	YGR192C_P	TDH3,GLD1, HSP35,HSP36, SSS2, GAPDH, GPD	involved in glycolysis and gluconeogenesis	54.89182825
S2_ENO2	YHR174W_P	ENO2	Expression induced in response to glucose	39.13934387
S3_HXK1	YFR053C_P	HXK1	a cytosolic protein that catalyzes phosphorylation of glucose during glucose metabolism;	19.88117423
S4_HSP12	YFL014W_P	HSP12,GLP1,HOR5	Plasma membrane protein involved in maintaining membrane organization and involved in maintaining organization during stress conditions	17.34087061
S5_OCH1	YGL038C_P	OCH1,LDB12,NGD29	initiates the polymannose outer chain elongation of N-linked oligosaccharides of glycoproteins	13.73445404

S6_STE2	YFL026W_P	STE2	Receptor for alpha-factor pheromone	11.8367168
S7_RPS27B	YHR021C_P	RPS27B	Protein component of the small (40S) ribosomal subunit	10.44802085
S8_ACT1	YFL039C_P	ACT1,ABY1,END7	Actin; structural protein involved in cell polarization, endocytosis, and other cytoskeletal functions	9.026818019
S9_PNC1	YGL037C_P	PNC1	Nicotinamidase that converts nicotinamide to nicotinic acid	7.733325771
S10_RPL22B	YFL034C-A_P	RPL22B,YFL035C-B	Ribosomal 60S subunit protein L22A;	7.673506581
S11_CRH1	YGR189C_P	CRH1	Chitin transglycosylase	7.485531483
S12_PAL2	YHR097C_P	PAL2	Putative protein of unknown function	5.947282823
S13_SOD2	YHR008C_P	SOD2	Mitochondrial manganese superoxide dismutase	5.77130881
S14_AGA2	YGL032C_P	AGA2	Adhesion subunit of a-agglutinin of a-cells	5.723434509
S15_HAC1	YFL031W_P	HAC1,ERN4,IRE15	Basic leucine zipper (bZIP) transcription factor (ATF/CREB1 homolog)	5.556340371
S16_PXR1	YGR280C_P	PXR1,GNO1	Essential protein involved in rRNA and snoRNA maturation	5.54263731
S17_SCW4	YGR279C_P	SCW4	Cell wall protein with similarity to glucanases	5.150137295
S18_QCR9	YGR183C_P	QCR9,UCR9	Subunit 9 of ubiquinol cytochrome-c reductase (Complex III)	5.030179087

S19_MPC2	YHR162W_P	MPC2	Highly conserved subunit of the mitochondrial pyruvate carrier (MPC)	4.904490091
S20_RIM101	YHL027W_P	RIM101,RIM1	Cys2His2 zinc-finger transcriptional repressor	4.867606497
S21_BGL2	YGR282C_P	BGL2	Endo-beta-1,3-glucanase	4.849359642
S22_RTC3	YHR087W_P	RTC3	Protein of unknown function involved in RNA metabolism	4.710959075
S23_FUR1	YHR128W_P	FUR1	High affinity uridine permease	4.625625701
S24_TNA1	YGR260W_P	TNA1	High affinity nicotinic acid plasma membrane permease	4.570500849
S25_SOL4	YGR248W_P	SOL4	6-phosphogluconolactonase	4.351104628
S26_TIM10	YHR005C-A_P	TIM10,MRS11	Essential protein of the mitochondrial intermembrane space	4.211253184

Table D.2 26 'moderate strength' promoters and the characterised activity from the Guo et al. (2015) publication. The corresponding ORFs and key functions as described in the Saccaromyces genome database (SGD - https://www.yeastgenome.org) of each of the 26 promoters in this library.

Promoter name	Promoter ID	ORF name and aliases	Key functions of ORF (From SGD)	Characterised promoter activity (Guo et al., 2015)
M1_AAD6	YFL056C_P	AAD6	Putative aryl-alcohol dehydrogenase; involved in oxidative stress response	4.104409456
M2_DIA4	YHR011W_P	DIA4	Probable mitochondrial seryl-tRNA synthetase	4.071287079
M3_GPA1	YHR005C_P	GPA1,CDC70, DAC1,SCG1	Subunit of the G protein involved in pheromone response	3.992346999
M4_YPT1	YFL038C_P	YPT1,YP2	Rab family GTPase; involved in the ER-to-Golgi step of the secretory pathway	3.970135572
M5_SBP1	YHL034C_P	SBP1,SSB1,SSBR1	Protein that binds eIF4G and has a role in repression of translation	3.883994347
M6_BCD1	YHR040W_P	BCD1	Essential protein required for the accumulation of box C/D snoRNA	3.795452557
M7_LSC2	YGR244C_P	LSC2	Beta subunit of succinyl-CoA ligase	3.736214818
M8_GRE3	YHR104W_P	GRE3	Aldose reductase; involved in methylglyoxal, d-xylose, arabinose, and galactose metabolism	3.62742987
M9_SBE22	YHR103W_P	SBE22	Protein involved in bud growth	3.588903078

M10_FSH1	YHR049W_P	FSH1	Putative serine hydrolase	3.469597717
M11_RPL2A	YFR031C-A_P	RPL2A,RPL5B	Ribosomal 60S subunit protein L2A	3.443688243
M12_ELP2	YGR200C_P	ELP2,KTI3,TOT2	Subunit of Elongator complex	3.438429882
M13_VMA10	YHR039C-A_P	VMA10, YHR039BC, YHR039C-B	part of the electrogenic proton pump found throughout the endomembrane system	3.395751964
M14_PAU13	YHL046C_P	PAU13	Protein of unknown function	3.305477958
M15_RPN1	YHR027C_P	RPN1,HRD2, NAS1	Non-ATPase base subunit of the 19S regulatory particle of the 26S proteasome	3.224045514
M16_UBP6	YFR010W_P	UBP6	Ubiquitin-specific protease	3.205051702
M17_SSZ1	YHR064C_P	SSZ1,PDR13	Hsp70 protein that interacts with Zuo1p (a DnaJ homolog)	3.170530869
M18_SLT2	YHR030C_P	SLT2,BYC2, MPK1,SLK2	Serine/threonine MAP kinase	3.128472831
M19_ERG1	YGR175C_P	ERG1	Squalene epoxidase	3.113875029
M20_NIC96	YFR002W_P	NIC96	Linker nucleoporin component of the nuclear pore complex (NPC)	3.088005687
M21_PRP8	YHR165C_P	PRP8,DBF 3,DNA39,RNA8, SLT21,USA2	Component of U4/U6-U5 snRNP complex	2.980500388
M22_ERG11	YHR007C_P	ERG11,CYP51	Lanosterol 14-alpha-demethylase	2.84351501

M23_GCN20	YFR009W_P	GCN20	Positive regulator of the Gcn2p kinase activity	2.837106469
M24_BAT1	YHR208W_P	BAT1,ECA39, TWT1	Mitochondrial branched-chain amino acid (BCAA) aminotransferase	2.808829008
M25_PUT2	YHR037W_P	PUT2	Delta-1-pyrroline-5-carboxylate dehydrogenase	2.77620287
M26_DED81	YHR019C_P	DED81	Cytosolic asparaginyl-tRNA synthetase	2.751568008

Table D.3 26 'moderate strength' promoters and the characterised activity from the Guo et al. (2015) publication. The corresponding ORFs and key functions as described in the Saccaromyces genome database (SGD - https://www.yeastgenome.org) of each of the 26 promoters in this library.

Promoter name	Promoter ID	ORF name and aliases	Key functions of ORF (From SGD)	Characterised promoter activity (Guo et al., 2015)
W1_YGR237C_P	YGR237C_P		Putative protein of unknown function	1.504220519
W2_OKP1	YGR179C_P	OKP1	Outer kinetochore protein required for accurate chromosome segregation;	1.485590731
W3_YPI1	YFR003C_P	YPI1	Regulatory subunit of the type I protein phosphatase (PP1) Glc7p	1.47778766
W4_SMX2	YFL017W-A_P	SMX2,SNP2, YFL018W-A	Core Sm protein Sm G	1.469776401
W5_PRE4	YFR050C_P	PRE4	Beta 7 subunit of the 20S proteasome	1.468914415
W6_HSN1	YHR127W_P	HSN1	Protein of unknown function	1.456243938
W7_MRP4	YHL004W_P	MRP4	Mitochondrial ribosomal protein of the small subunit	1.44218121
W8_YHL005C_P	YHL005C_P		Putative protein of unknown function	1.411164425
W9_ YGR161W-C_P	YGR161W-C_P		Putative protein of unknown function	1.399272663
W10_SET5	YHR207C_P	SET5	Methyltransferase involved in methylation of histone H4 Lys5, -8, -12	1.396795209

W11_LSM12	YHR121W_P	LSM12	Protein of unknown function that may function in RNA processing	1.394645283
W12_RNA15	YGL044C_P	RNA15	Component of the cleavage and polyadenylation factor I (CF I);	1.393247254
W13_RBG2	YGR173W_P	RBG2,GIR1	Protein with a role in translation	1.358739807
W14_OPI1	YHL020C_P	OPI1	Transcriptional regulator of a variety of genes;	1.345962226
W15_GTR2	YGR163W_P	GTR2	Subunit of a TORC1-stimulating GTPase and the EGO/GSE complex	1.345725653
W16_RIM15	YFL033C_P	RIM15,TAK1	Protein kinase involved in cell proliferation in response to nutrients	1.338333155
W17_GAR1	YHR089C_P	GAR1	Protein component of the H/ACA snoRNP pseudouridylase complex	1.326412231
W18_DST1	YGL043W_P	DST1,PPR2	General transcription elongation factor TFIIS	1.325558405
W19_CGR1	YGL029W_P	CGR1	Protein involved in nucleolar integrity and processing of pre-rRNA	1.319417087
W20_TFG1	YGR186W_P	TFG1,RAP74, SSU71	TFIIF (Transcription Factor II) largest subunit;	1.304559609
W21_ YGR164W_P	YGR164W_P		Putative protein of unknown function	1.302921043
W22_ YHR138C_P	YHR138C_P		Protein of unknown function	1.294836568
W23_ATE1	YGL017W_P	ATE1	Arginyl-tRNA-protein transferase	1.289141123

W24_NMD3	YHR170W_P	NMD3,SRC5	Protein involved in nuclear export of the large ribosomal subunit	1.281306985
W25_ YFR032C-B_P	YFR032C-B_P		Putative protein of unknown function	1.277211553
W26_DAP2	YHR028C_P	DAP2,DPP2	Dipeptidyl aminopeptidase	1.269440802

Appendix E

Supplementary figures and tables - for Chapter 5

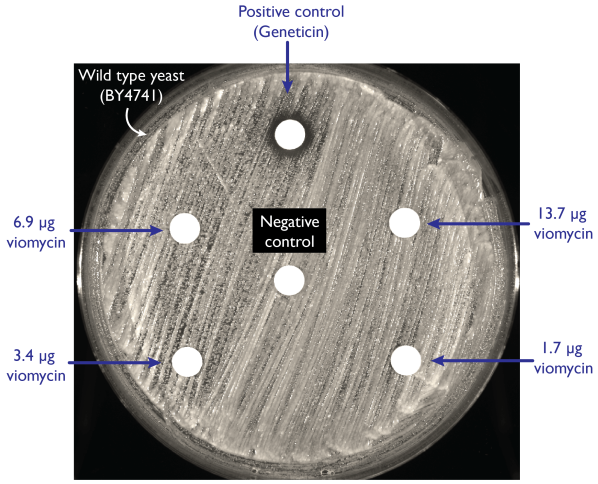


Figure E.1

A disk diffusion assay screening wild type yeast (BY4741) for susceptibility to viomycin

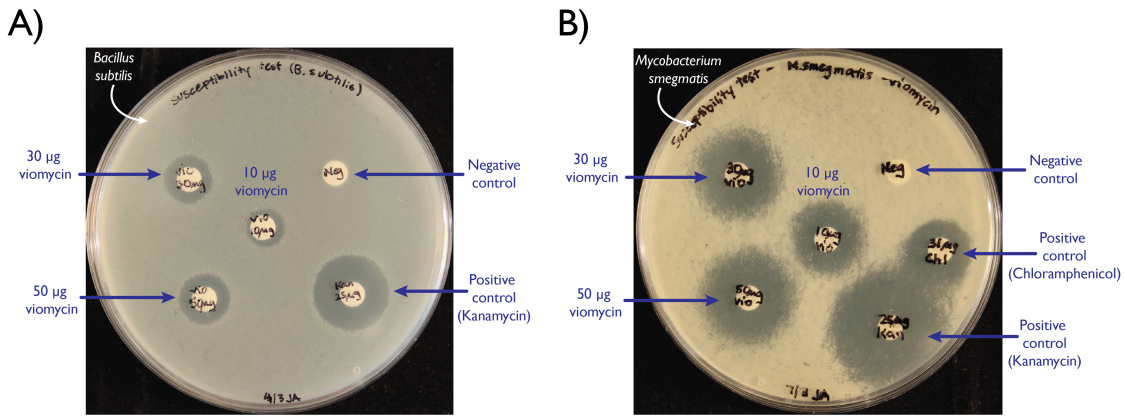
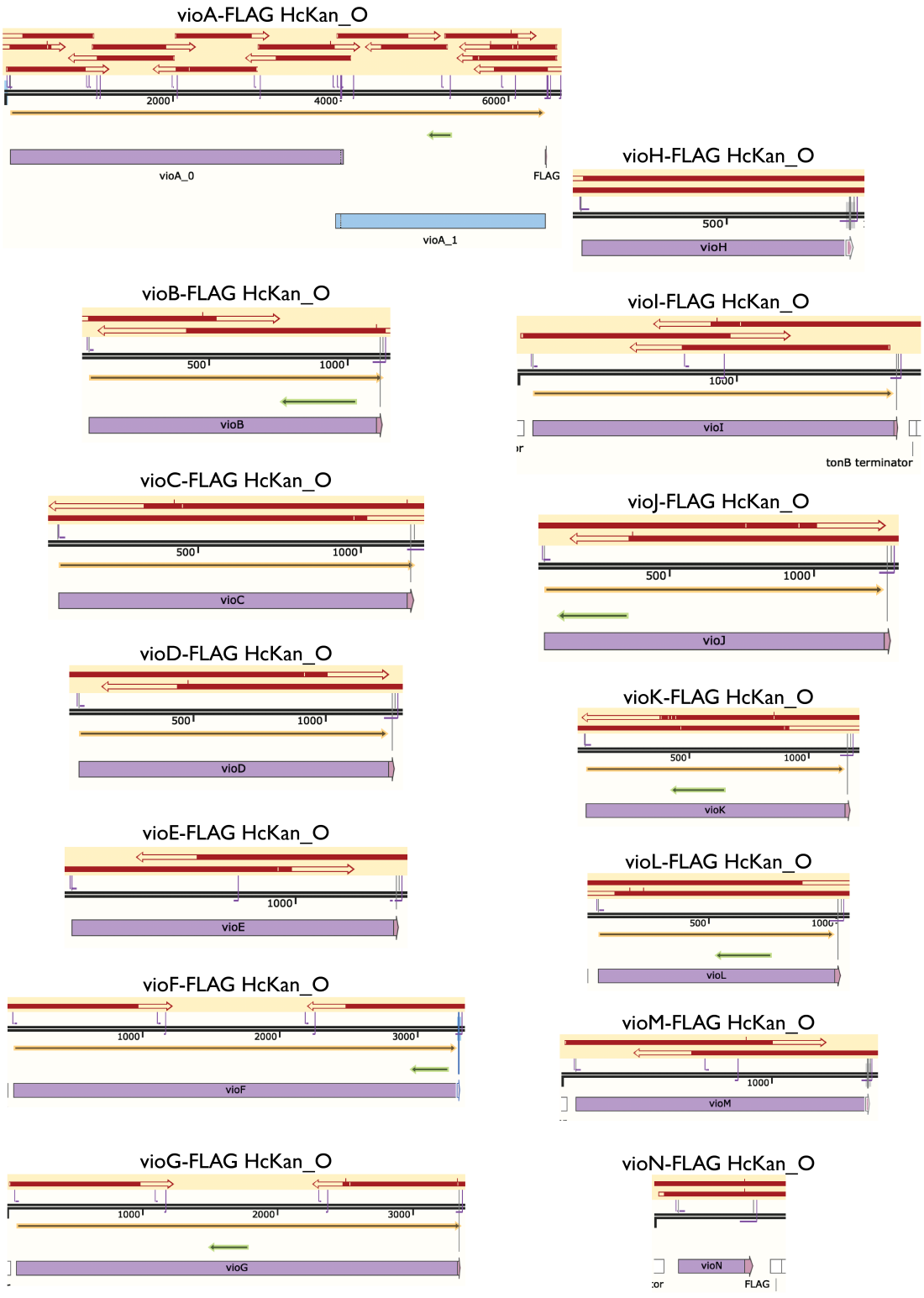


Figure E.2

A disk diffusion assay screening *B. subtilis* (left) and *M. smegmatis* (right) for their susceptibility to different concentrations of viomycin


vioA-FLAG HcKan_O
2000
4000
6000
vioA_0
FLAG
vioA_1
vioH-FLAG HcKan_O
500
vioH
vioB-FLAG HcKan_O
500
1000
vioB
vioI-FLAG HcKan_O
1000
vioI
tonB terminator
vioC-FLAG HcKan_O
500
1000
vioC
vioJ-FLAG HcKan_O
500
1000
vioJ
vioD-FLAG HcKan_O
500
1000
vioD
vioK-FLAG HcKan_O
500
1000
vioK
vioE-FLAG HcKan_O
1000
vioE
vioL-FLAG HcKan_O
500
1000
vioL
vioF-FLAG HcKan_O
1000
2000
3000
vioF
vioM-FLAG HcKan_O
1000
vioM
vioG-FLAG HcKan_O
1000
2000
3000
vioG
vioN-FLAG HcKan_O
vioN
FLAG

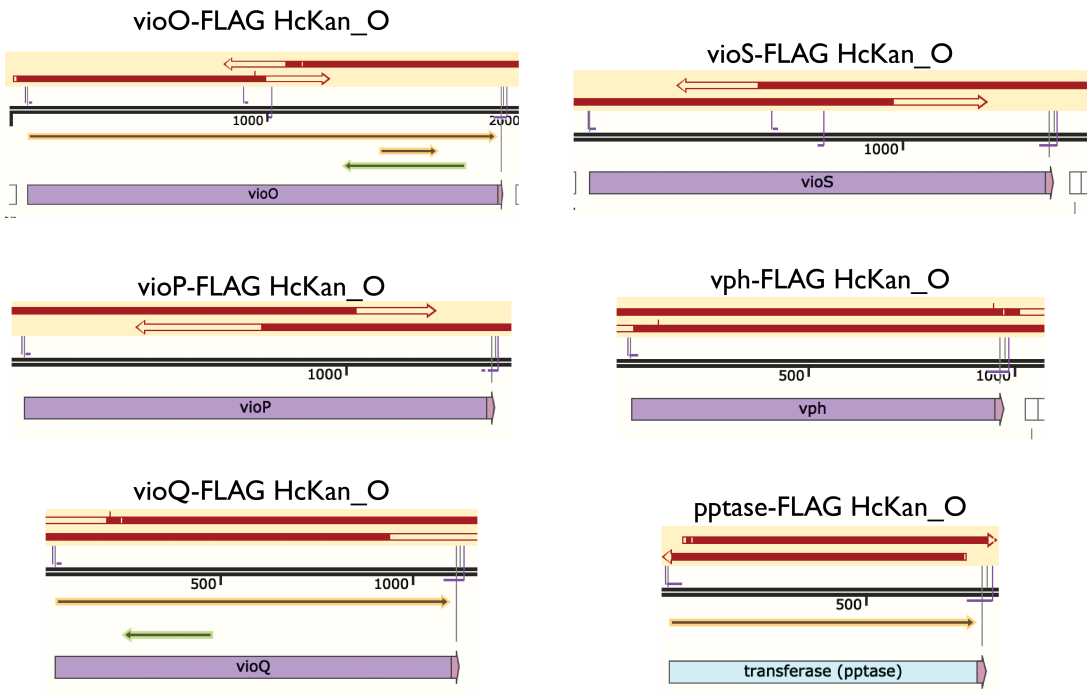


Figure E.3

Sequencing verification results for all 20 viomycin BSGs domesticated to the YeastFab format

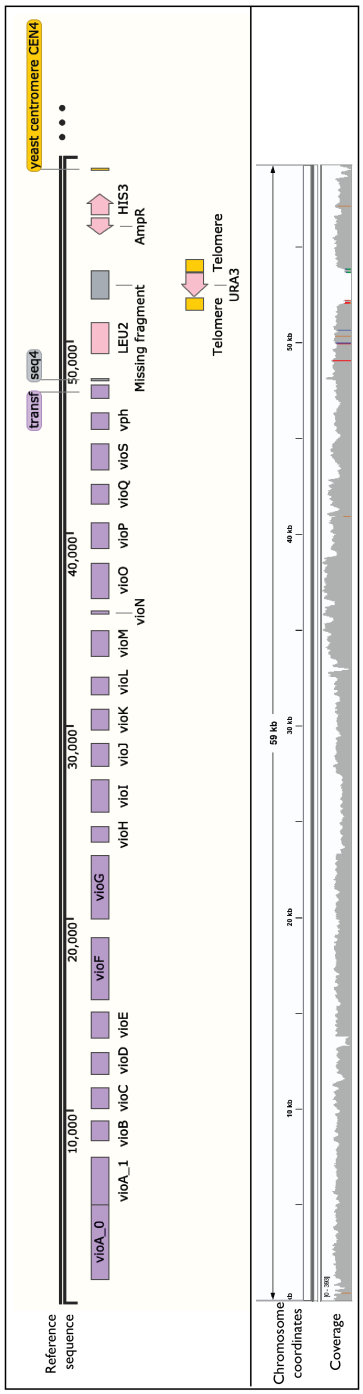


Figure E.4

On the left is the reference sequence for the YAC containing the viomycin pathway visualised using SnapGene. On the right is the sequencing alignment of next generation sequencing reads visualised using IGV, showing the read coverage at each segment of the chromosome. Sequencing of the sample was carried out at the Sanger Institute at Cambridge. Analysis of the raw sequencing results was carried out by Dr. Daniel Schindler and Marco Monti

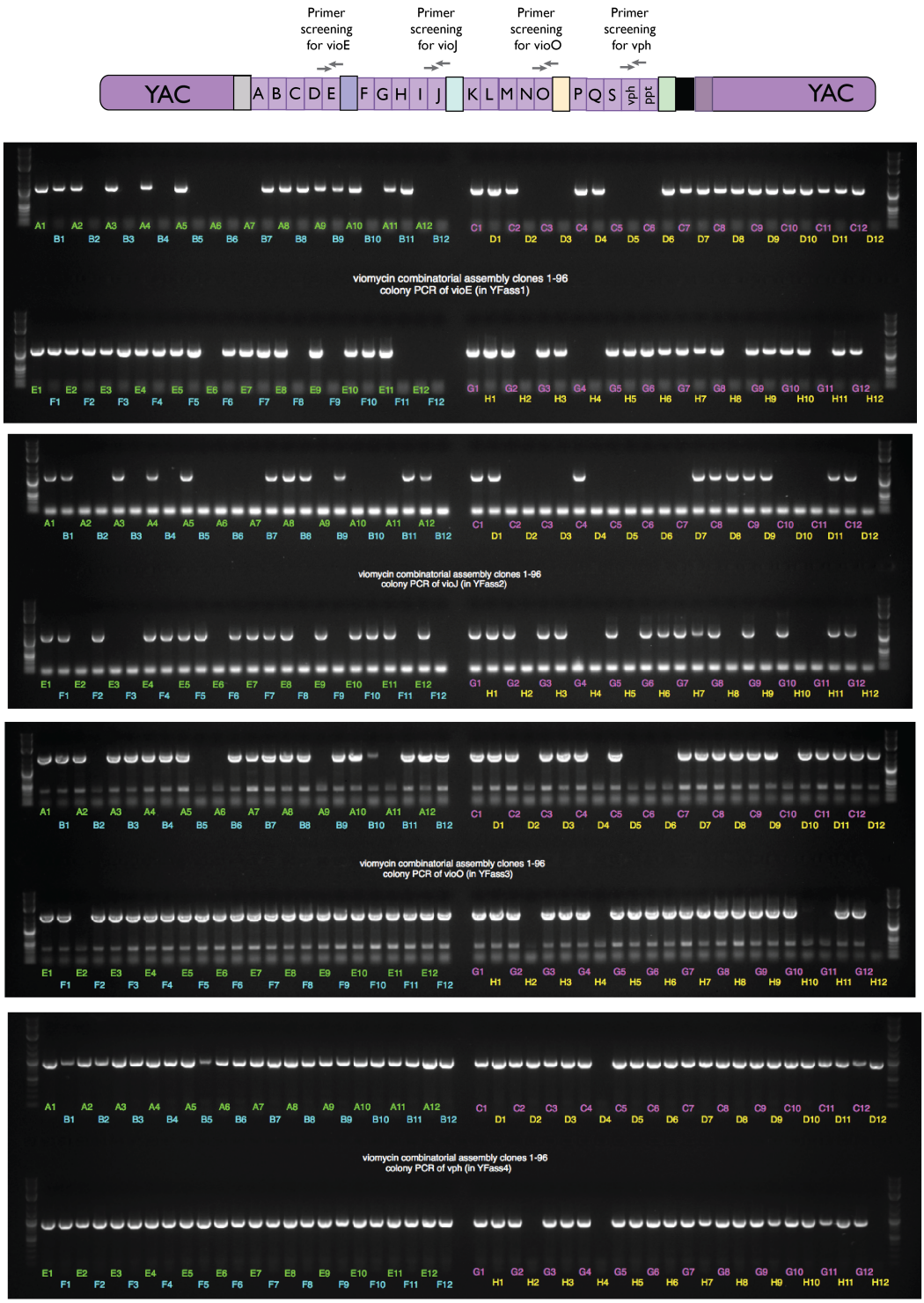


Figure E.5

Diagrammatic representation of primer binding sites in the YAC and the colony PCR verification results for all 96 clones.

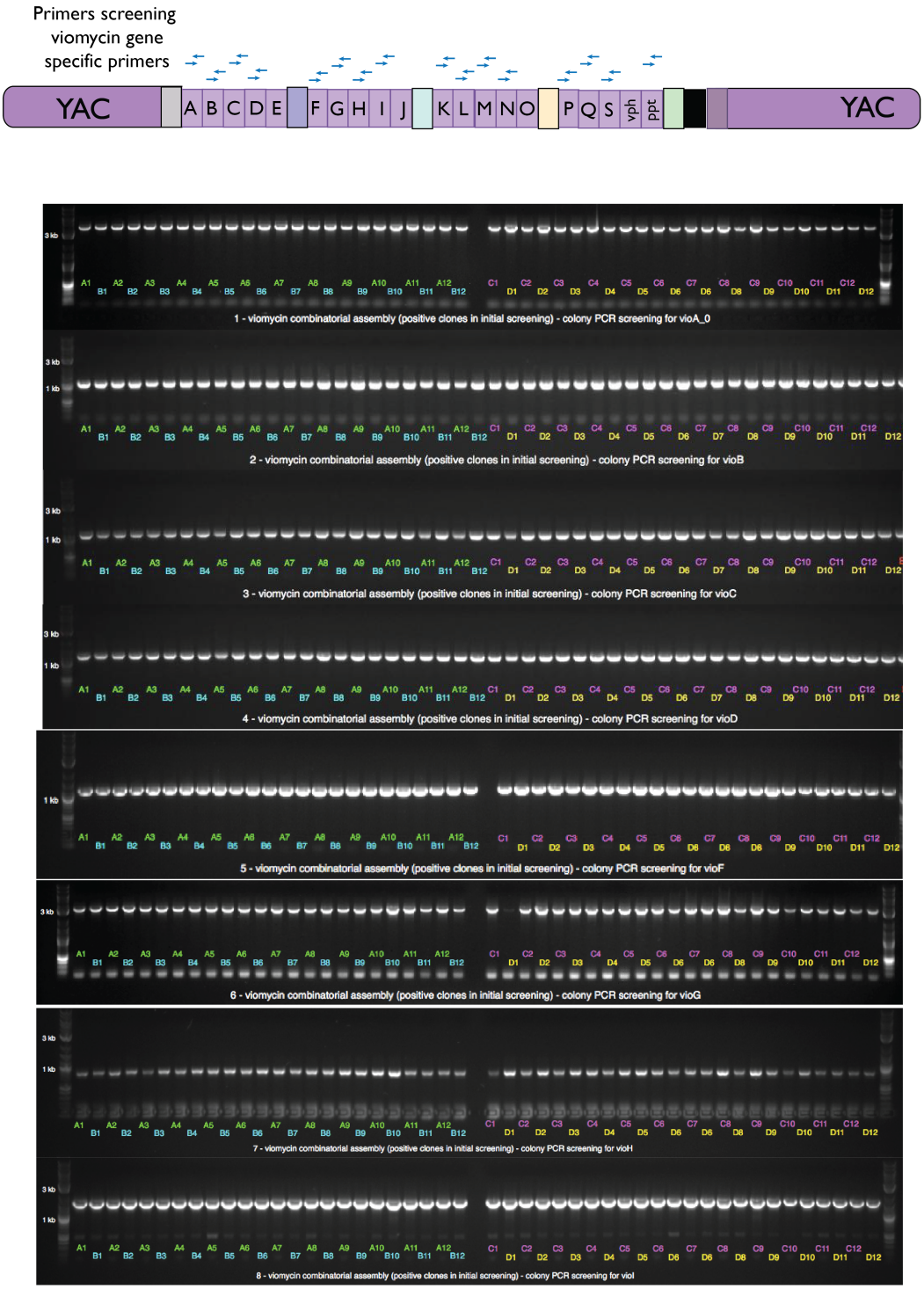
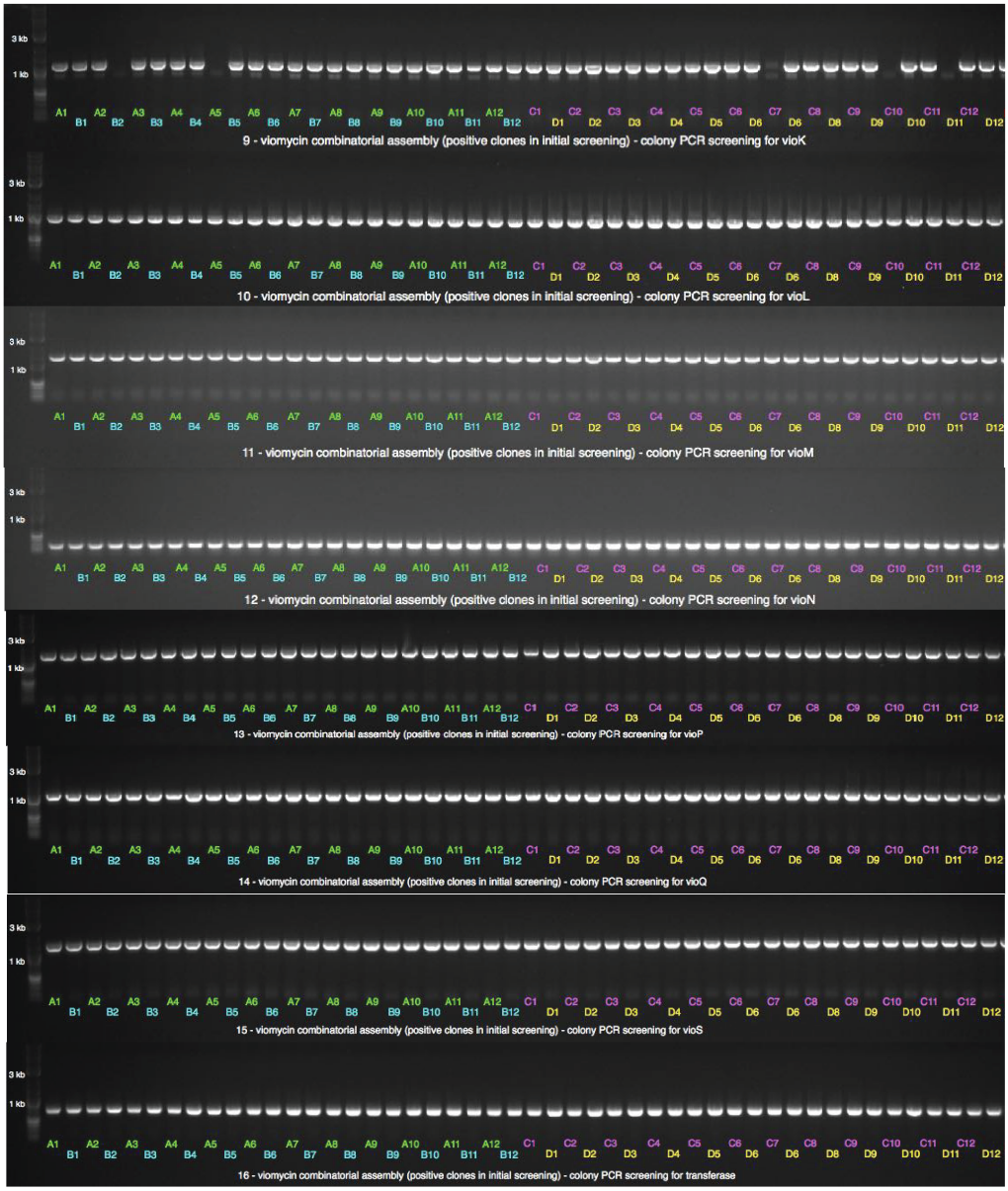


Figure E.6

Diagrammatic representation of primers used for screening the YAC and the colony PCR verification results for all 48 clones.


3 kb
1 kb
A1 B1 A2 B2 A3 B3 A4 B4 A5 B5 A6 B6 A7 B7 A8 B8 A9 B9 A10 B10 A11 B11 A12 B12 C1 D1 C2 D2 C3 D3 C4 D4 C5 D5 C6 D6 C7 D6 C8 D8 C9 D9 C10 D10 C11 D11 C12 D12
9 - viomycin combinatorial assembly (positive clones in initial screening) - colony PCR screening for vioK
10 - viomycin combinatorial assembly (positive clones in initial screening) - colony PCR screening for vioL
11 - viomycin combinatorial assembly (positive clones in initial screening) - colony PCR screening for vioM
12 - viomycin combinatorial assembly (positive clones in initial screening) - colony PCR screening for vioN
13 - viomycin combinatorial assembly (positive clones in initial screening) - colony PCR screening for vioP
14 - viomycin combinatorial assembly (positive clones in initial screening) - colony PCR screening for vioQ
15 - viomycin combinatorial assembly (positive clones in initial screening) - colony PCR screening for vioS
16 - viomycin combinatorial assembly (positive clones in initial screening) - colony PCR screening for transferase

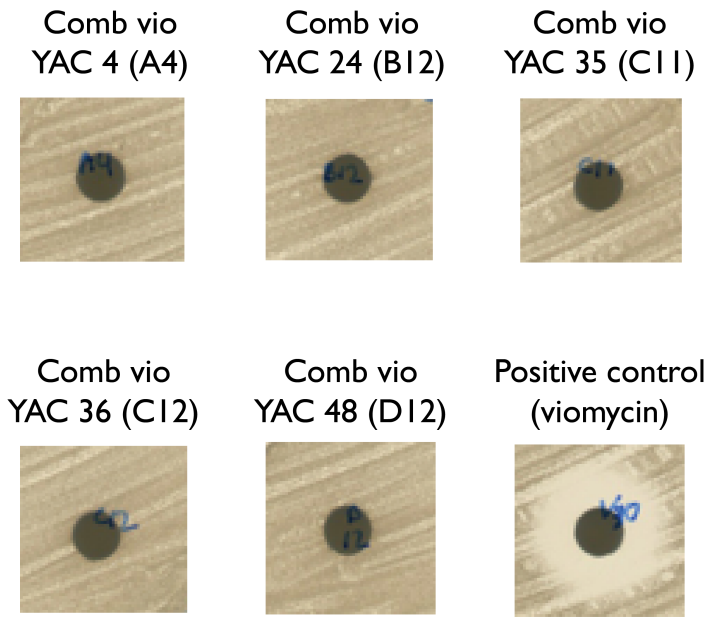


Figure E.7

Screening of 5 clones (Comb YAC4, Comb YAC24, Comb YAC35, Comb YAC36, Comb YAC48) from combinatorial viomycin assembly by disk diffusion assay. This work was carried out at the Wright lab (McMaster University) by Beth Culp, Dr. Grace Yim and Dr. Venkateswarlu Yarlagadda

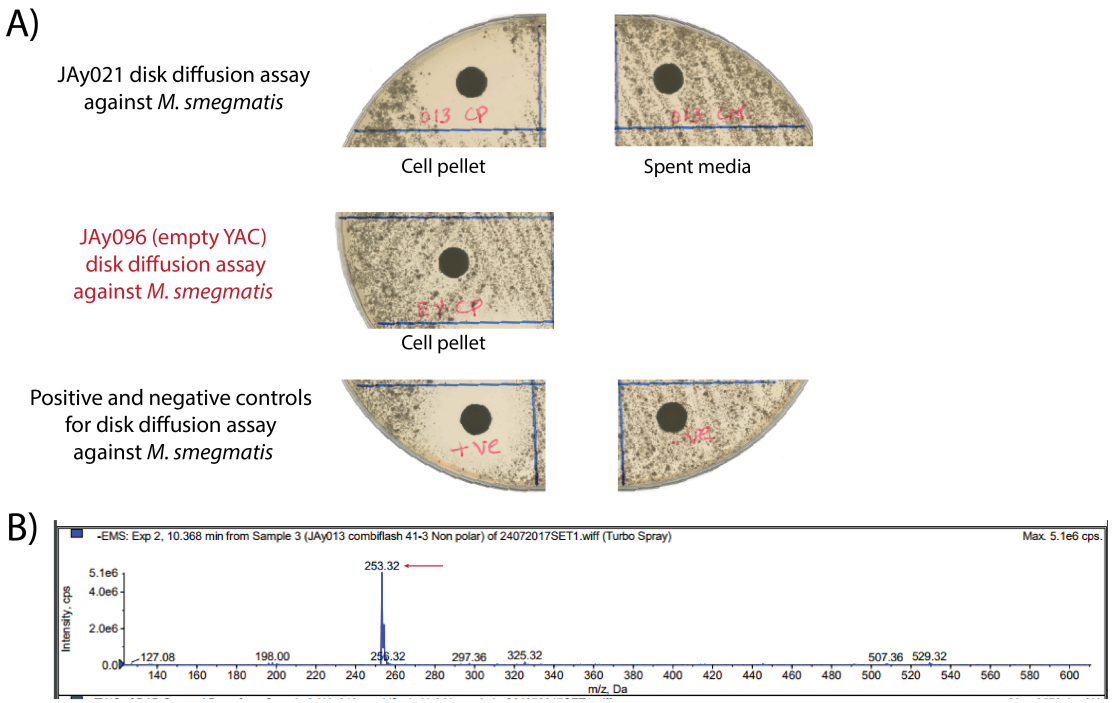


Figure E.8

A) Disk diffusion assay of yeast extracts of JAy021 and JAy096 (negative control) against *M. smegmatis*. B) Chemical analysis of JAy021 cell pellet extract by LC-MS and the identification of a peak with a mass to charge ratio of 253.32. This work was carried out at the Wright lab (McMaster University) by Beth Culp, Dr. Grace Yim and Dr. Venkateswarlu Yarlagadda

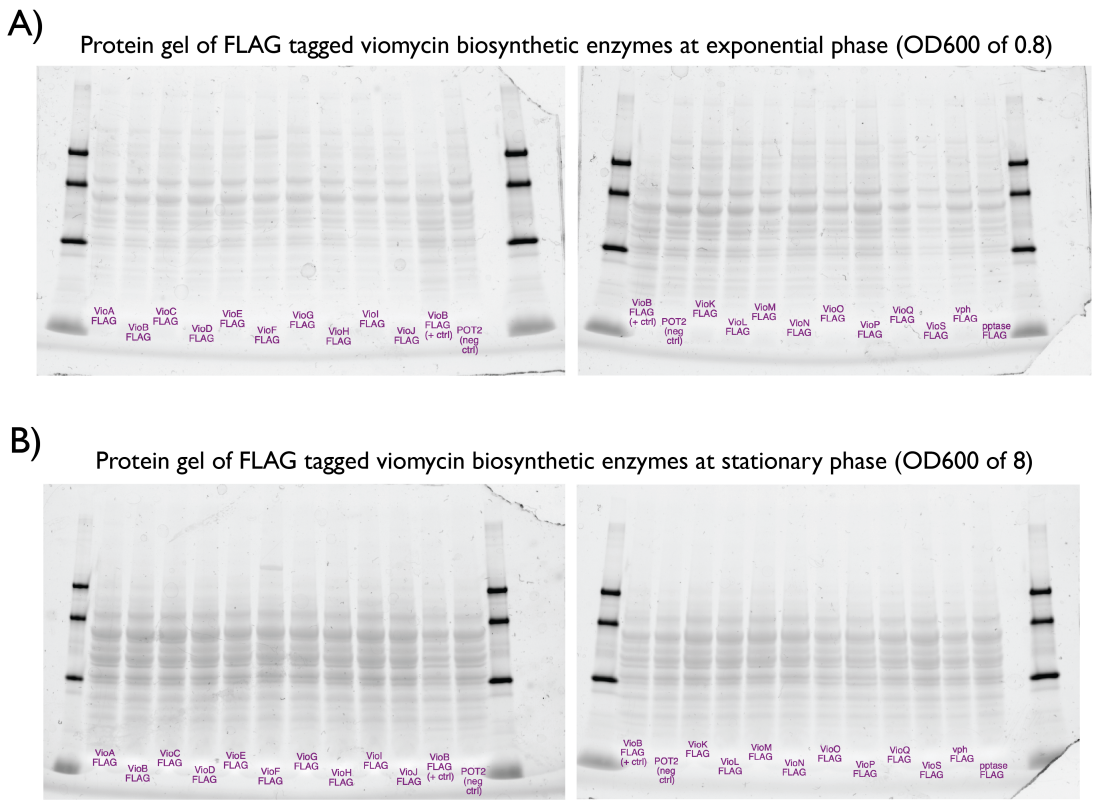


Figure E.9

Stain free gel showing the total protein of yeast cell lysates in extracts taken at A) exponential phase and B) stationary phase.

Appendix F

Supplementary figures and tables - for Chapter 6

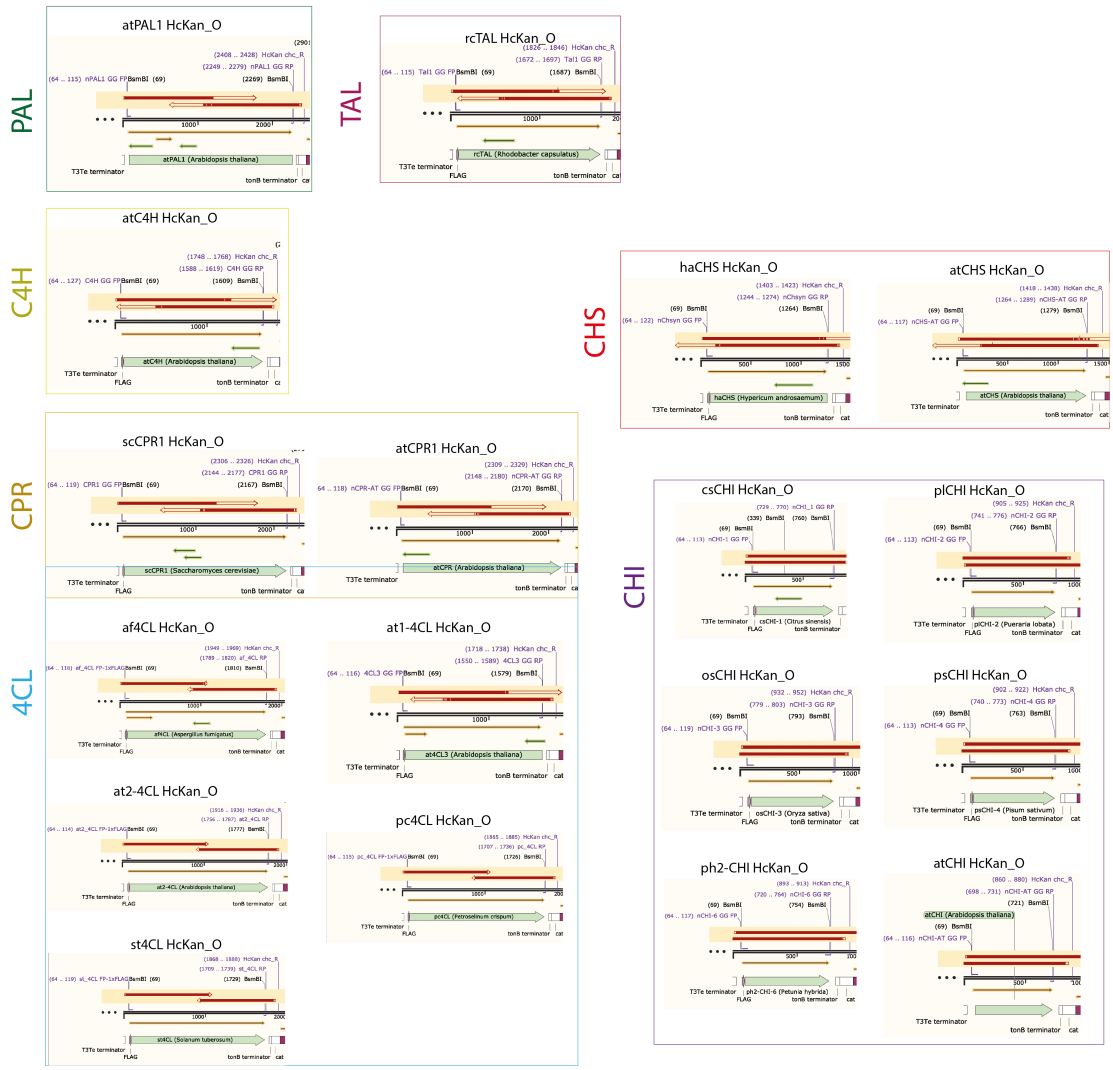


Figure F.1

Sequencing verification results of naringenin BSGs (PAL, TAL, C4H, CPR, 4CL) in HcKan_O vector

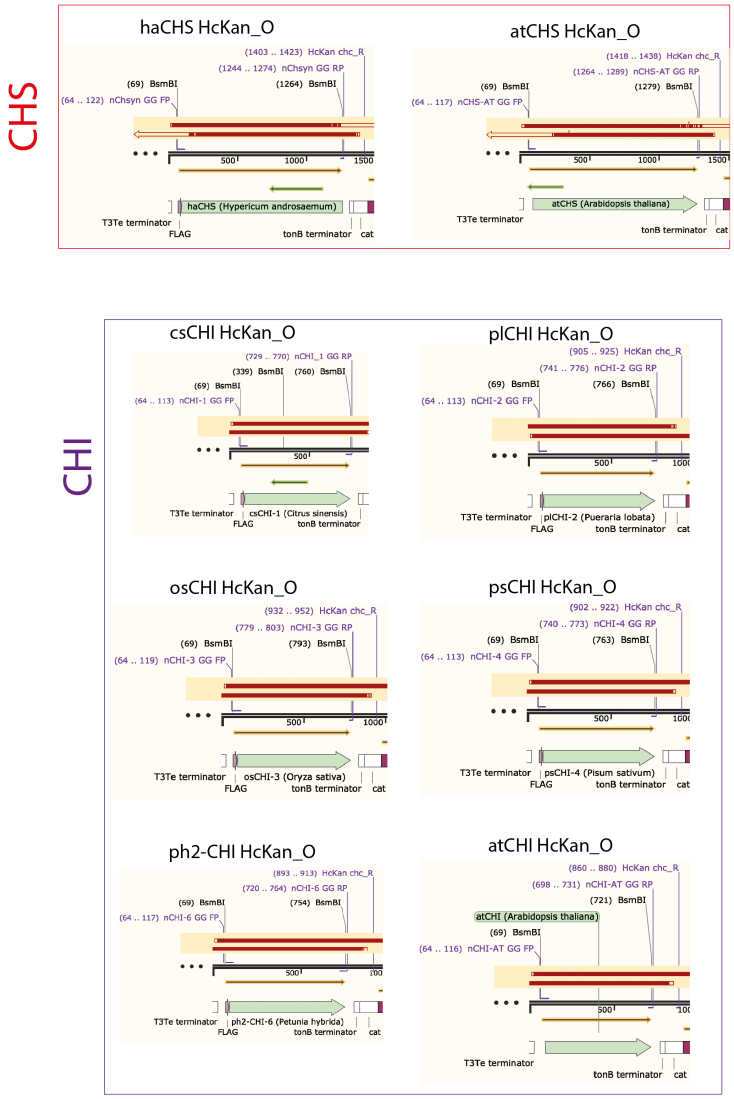


Figure F.2

Sequencing verification results of naringenin BSGs (CHS and CHI) in HcKan_O vector

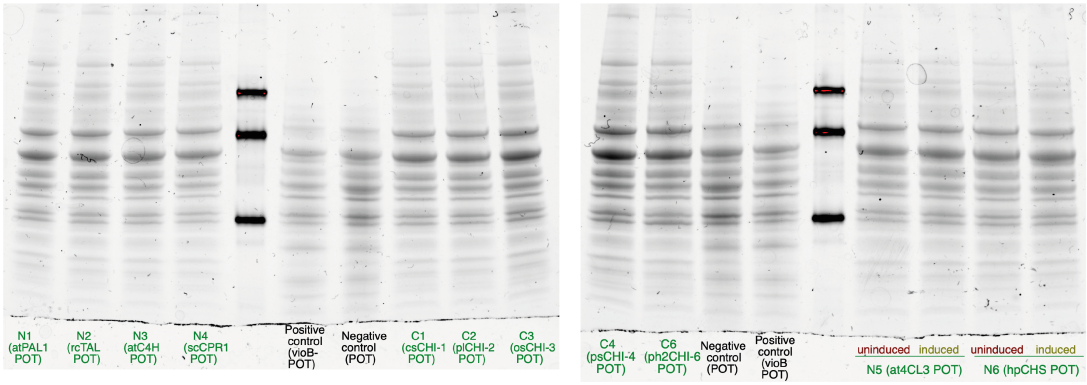


Figure F.3

Stain free gel showing the total protein of yeast cell lysates containing naringenin POTs

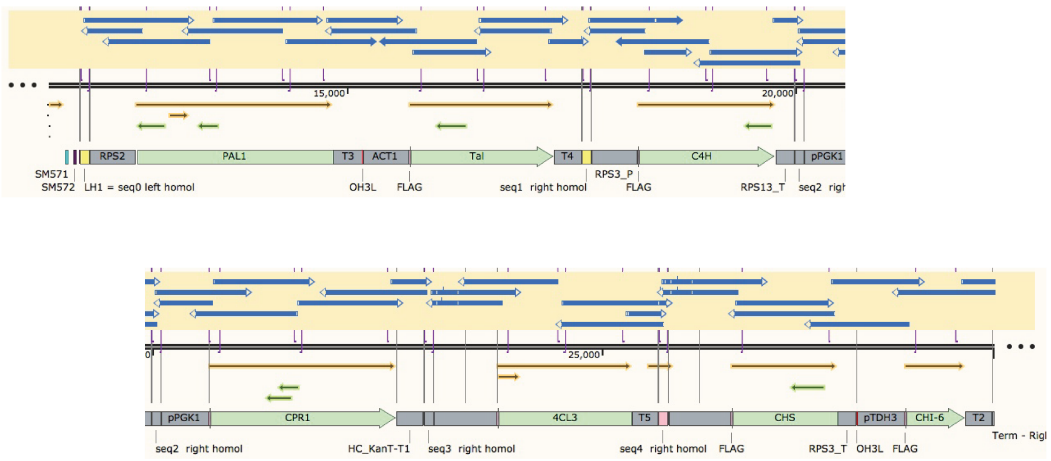


Figure F.4

Sequencing verification of JAy081

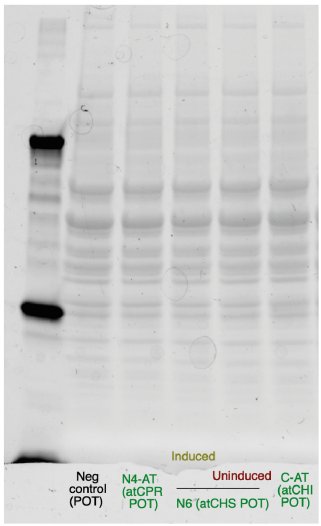


Figure F.5

Stain free gel showing the total protein of yeast cell lysates containing A. thaliana naringenin POTs

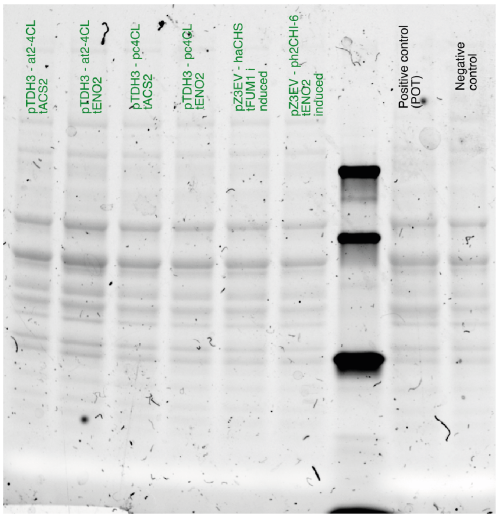


Figure F.6

Stain free gel showing the total protein of yeast cell lysates containing naringenin POTs for the truncated naringenin pathway

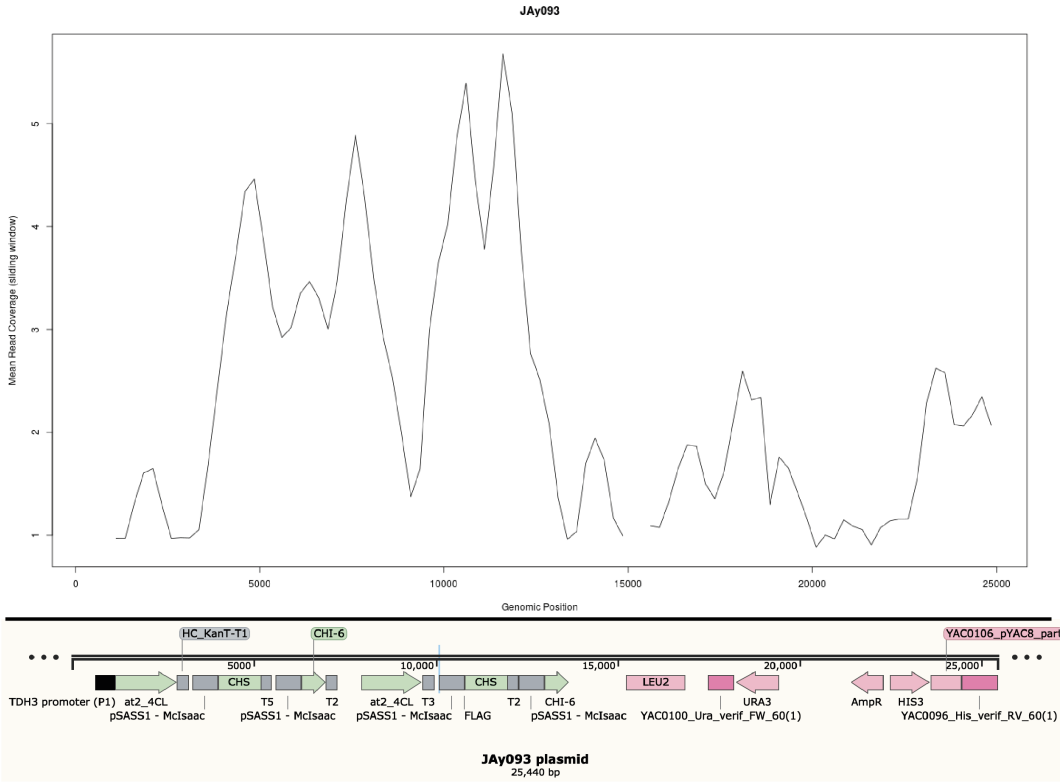


Figure F.7

The top panel shows the mean read coverage for the plasmid JAy093 by nanopore sequencing. The bottom panel shows the plasmid map for JAy093 and the location of the naringenin BSGs.

BSG name	BSG function	BSG enzyme description	Origin species	Genbank accession number	Protein ID
atPAL	PAL1	phenylalanine ammonia-lyase 1	*Arabidopsis thaliana (thale cress)*	15228074	N/A
rcTAL	TAL	tyrosine ammonia-lyase	*Rhodobacter capsulatus*	RCAP_rcc01072	D5AR26
atC4H	C4H	cinnamate 4-hydroxylase	*Arabidopsis thaliana*	U71081	AAB58356.1
scCPR	CPR1	NADPH-cytochrome P450 reductase CPR1	*Saccharomyces cerevisiae*	D13788	BAA02936.1
atCPR		cytochrome P450 reductase ATR1	*Arabidopsis thaliana*	NM_118585	NP_194183.1
at1-4CL	4CL3	4-coumarate-CoA ligase 3	*Arabidopsis thaliana (thale cress)*	30697142	N/A
nt4CL		4-coumarate-CoA ligase	*Nicotiana tabacum*	U50845	AAB18637.1
pc4CL		4-coumarate-CoA ligase	*Petroselinum crispum*	X13325	CAA31697.1
at2-4CL		4-coumarate-CoA ligase	*Arabidopsis thaliana*	U18675	AAA82888.1
st4CL		4-coumarate-CoA ligase	*Solanum tuberosum*	M62755	AAA33842.1
af4CL		4-coumarate-CoA ligase	*Aspergillus fumigatus*	BX649606	CAF32026.1
haCHS	CHS	chalcone synthase	*Hypericum androsaemum*	AF315345	AAG30295.1
atCHS		chalcone synthase	*Arabidopsis thaliana (thale cress)*	15240753	N/A
csCHI	CHI	chalcone isomerase	*Citrus sinensis*	AB011794	BAA36552.1
plCHI		chalcone isomerase	*Pueraria lobata*	D63577	BAA09795.1
osCHI		chalcone isomerase	*Oryza sativa*	NM_001058249	NP_001051714.1
psCHI		chalcone isomerase	*Pisum sativum*	U03433	AAA50174.1
ph1-CHI		chalcone isomerase	*Petunia hybrida*	X14589	CAA32729.1

Table F.1 continued from previous page

BSG name	BSG function	BSG enzyme description	Origin species	Genbank accession number	Protein ID
ph2-CHI		chalcone isomerase	*Petunia hybrida*	X14590	CAA32730.1
atCHI		chalcone isomerase	*Arabidopsis thaliana (thale cress)*	18414838	N/A

Bibliography

Abatemarco, J., Sarhan, M. F., Wagner, J. M., Lin, J.-L., Liu, L., Hassouneh, W., Yuan, S.-F., Alper, H. S., and Abate, A. R. (2017). RNA-aptamers-in-droplets (RAPID) high-throughput screening for secretory phenotypes. *Nature Communications*, 8(1):332. (page 196)

Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S., Stracquadanio, G., Richardson, S. M., Dymond, J. S., Kuang, Z., Scheifele, L. Z., Cooper, E. M., et al. (2014). Total synthesis of a functional designer eukaryotic chromosome. *Science*, 344(6179):55–58. (page 190)

Apel, A. R., D'Espaux, L., Wehrs, M., Sachs, D., Li, R. A., Tong, G. J., Garber, M., Nnadi, O., Zhuang, W., Hillson, N. J., Keasling, J. D., and Mukhopadhyay, A. (2017). A Cas9-based toolkit to program gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Research*. (pages 58, 61, 66, 67, 71, 99, and 100)

Asadollahi, M. A., Maury, J., Patil, K. R., Schalk, M., Clark, A., and Nielsen, J. (2009). Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through in silico driven metabolic engineering. *Metabolic Engineering*, 11(6):328–334. (page 6)

Balleza, E., Kim, J. M., and Cluzel, P. (2018). Systematic characterization of maturation time of fluorescent proteins in living cells. *Nature Methods*, 15(1):47. (page 50)

Bao, Z., HamediRad, M., Xue, P., Xiao, H., Tasan, I., Chao, R., Liang, J., and Zhao, H. (2018). Genome-scale engineering of *Saccharomyces cerevisiae* with single-nucleotide precision. *Nature Biotechnology*, 36(6):505–508. (page 195)

Barbieri, E. M., Muir, P., Akhuetie-Oni, B. O., Yellman, C. M., and Isaacs, F. J. (2017). Precise editing at DNA replication forks enables multiplex genome engineering in eukaryotes. *Cell*, 171(6):1453–1467. (page 195)

Barkei, J. J., Kevany, B. M., Felnagle, E. A., and Thomas, M. G. (2009). Investigations into viomycin biosynthesis by using heterologous production in *Streptomyces lividans*. *Chembiochem*, 10(2):366–376. (page 108)

Blazeck, J. and Alper, H. S. (2013). Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnology journal*, 8(1):46–58. (page 59)

Blount, B., Gowers, G. F., Ho, J., Ledesma-Amaro, R., Jovicevic, D., McKiernan, R., Xie, Z., Li, B., Yuan, Y., and Ellis, T. (2018). Rapid host strain improvement by *in vivo* rearrangement of a synthetic yeast chromosome. *Nature Communications*, 9(1):1932. (pages 190, 194, and 195)

Boussebayle, A., Torka, D., Ollivaud, S., Braun, J., Bofill-Bosch, C., Dombrowski, M., Groher, F., Hamacher, K., and Suess, B. (2019). Next-level riboswitch development—implementation of Capture-SELEX facilitates identification of a new synthetic riboswitch. *Nucleic Acids Research*, 47(9):4883–4895. (page 65)

Brindle, P., Holland, J., Willett, C., Innis, M., and Holland, M. (1990). Multiple factors bind the upstream activation sites of the yeast enolase genes ENO1 and ENO2: ABFI protein, like repressor activator protein RAP1, binds *cis*-acting sequences which modulate repression or activation of transcription. *Molecular and Cellular Biology*, 10(9):4872–4885. (pages 58 and 59)

Brochado, A. R., Matos, C., Møller, B. L., Hansen, J., Mortensen, U. H., and Patil, K. R. (2010). Improved vanillin production in baker's yeast through *in silico* design. *Microbial Cell Factories*, 9(1):84. (page 6)

Carmen, A. A., Brindle, P. K., Park, C. S., and Holland, M. J. (1995). Transcriptional regulation by an upstream repression sequence from the yeast enolase gene ENO1. *Yeast*, 11(11):1031–1043. (pages 58 and 59)

Copp, J. and Neilan, B. (2006). The phosphopantetheinyl transferase superfamily: phylogenetic analysis and functional implications in cyanobacteria. *Appl. Environ. Microbiol.*, 72(4):2298–2305. (page 141)

Cranfill, P. J., Sell, B. R., Baird, M. A., Allen, J. R., Lavagnino, Z., De Gruiter, H. M., Kremers, G.-J., Davidson, M. W., Ustione, A., and Piston, D. W. (2016). Quantitative assessment of fluorescent proteins. *Nature Methods*, 13(7):557. (page 180)

Critchlow, S. E., O'Dea, M. H., Howells, A. J., Couturier, M., Gellert, M., and Maxwell, A. (1997). The interaction of the f plasmid killer protein, Ccdb, with DNA gyrase: induction of DNA cleavage and blocking of transcription. *Journal of Molecular Biology*, 273(4):826–839. (page 51)

Da Silva, N. A. and Srikrishnan, S. (2012). Introduction and expression of genes for metabolic engineering applications in *Saccharomyces cerevisiae*. *FEMS Yeast Research*, 12(2):197–214. (page 60)

Dao-Thi, M.-H., Van Melderen, L., De Genst, E., Afif, H., Buts, L., Wyns, L., and Loris, R. (2005). Molecular basis of gyrase poisoning by the addiction toxin Ccdb. *Journal of Molecular Biology*, 348(5):1091–1102. (page 51)

Dashti, Y., Grkovic, T., and Quinn, R. J. (2014). Predicting natural product value, an exploration of anti-TB drug space. *Natural Product Reports*, 31(8):990–998. (page 103)

Decoene, T., De Paepe, B., Maertens, J., Coussement, P., Peters, G., De Maeseneire, S. L., and De Mey, M. (2018). Standardization in synthetic biology: an engineering discipline coming of age. *Critical Reviews in Biotechnology*, 38(5):647–656. (page 3)

Denis, C. L., Ferguson, J., and Young, E. (1983). mRNA levels for the fermentative alcohol dehydrogenase of *Saccharomyces cerevisiae* decrease upon growth on a nonfermentable carbon source. *Journal of Biological Chemistry*, 258(2):1165–1171. (page 60)

Dvir, S., Velten, L., Sharon, E., Zeevi, D., Carey, L. B., Weinberger, A., and Segal, E. (2013). Deciphering the rules by which 5-UTR sequences affect protein expression in yeast. *Proceedings of the National Academy of Sciences*, 110(30):E2792–E2801. (pages 54, 55, and 192)

Dymond, J. S., Richardson, S. M., Coombes, C. E., Babatz, T., Muller, H., Annaluru, N., Blake, W. J., Schwerzmann, J. W., Dai, J., Lindstrom, D. L., et al. (2011). Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature*, 477(7365):471. (page 190)

Ellington, A. D. and Szostak, J. W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature*, 346(6287):818. (page 65)

Endy, D. (2005). Foundations for engineering biology. *Nature*, 438(7067):449–453. (page 33)

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE*, 3(11). (page 33)

Frantz, S. (2004). Better antibiotics through chemistry. *Nature Reviews Drug Discovery*, 3:900–901. (page 104)

Garcia-Ruiz, E., Auxillos, J., Li, T., Dai, J., and Cai, Y. (2018). Yeastfab: High-throughput genetic parts construction, measurement, and pathway engineering in yeast. *Methods Enzymol*, 608:277–306. (page 41)

Giniger, E., Varnum, S. M., and Ptashne, M. (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell*, 40(4):767–774. (page 63)

Gould, K. (2016). Antibiotics: from prehistory to the present day. *Journal of Antimicrobial Chemotherapy*, 71(3):572–575. (page 103)

Groher, F., Bofill-Bosch, C., Schneider, C., Braun, J., Jager, S., Geißler, K., Hamacher, K., and Suess, B. (2018). Riboswitching with ciprofloxacin—development and characterization of a novel RNA regulator. *Nucleic Acids Research*, 46(4):2121–2132. (page 97)

Groher, F. and Suess, B. (2014). Synthetic riboswitches - A tool comes of age. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1839(10):964–973. (pages 65 and 97)

Guarente, L., Lalonde, B., Gifford, P., and Alani, E. (1984). Distinctly regulated tandem upstream activation sites mediate catabolite repression of the CYC1 gene of *S. cerevisiae*. *Cell*, 36(2):503–511. (pages 58, 59, and 60)

Guo, Y., Dong, J., Zhou, T., Auxillos, J., Li, T., Zhang, W., Wang, L., Shen, Y., Luo, Y., Zheng, Y., et al. (2015). Yeastfab: the design and construction of standard biological parts for metabolic engineering in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 43(13):e88–e88. (pages 34, 35, 42, 61, 66, 69, 73, 77, 99, and 107)

Hanson, S., Berthelot, K., Fink, B., McCarthy, J. E., and Suess, B. (2003). Tetracycline-aptamer-mediated translational regulation in yeast. *Molecular Microbiology*, 49(6):1627–1637. (page 66)

Hirai, H., Tani, T., and Kikyo, N. (2010). Structure and functions of powerful transactivators: VP16, MyoD and FoxA. *The International journal of developmental biology*, 54(11-12):1589. (page 85)

Holm, M., Borg, A., Ehrenberg, M., and Sanyal, S. (2016). Molecular mechanism of viomycin inhibition of peptide elongation in bacteria. *Proceedings of the National Academy of Sciences*, 113(4):978–983. (page 107)

Kanigowska, P., Shen, Y., Zheng, Y., Rosser, S., and Cai, Y. (2016). Smart DNA fabrication using sound waves: applying acoustic dispensing technologies to synthetic biology. *Journal of Laboratory Automation*, 21(1):49–56. (page 193)

Koch, M., Pandi, A., Borkowski, O., Batista, A. C., and Faulon, J.-L. (2019). Custom-made transcriptional biosensors for metabolic engineering. *Current Opinion in Biotechnology*, 59:78–84. (page 195)

Koopman, F., Beekwilder, J., Crimi, B., van Houwelingen, A., Hall, R. D., Bosch, D., van Maris, A. J., Pronk, J. T., and Daran, J.-M. (2012). *De novo* production of the flavonoid naringenin in engineered *Saccharomyces cerevisiae*. *Microbial Cell Factories*, 11(1):155. (pages 144, 145, 146, 163, 172, 178, and 193)

Lee, M. E., Aswani, A., Han, A. S., Tomlin, C. J., and Dueber, J. E. (2013). Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay. *Nucleic Acids Research*, 41(22):10668–10678. (page 71)

Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015). A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synthetic Biology*, 4(9):975–986. (page 61)

Lewis, K. (2013). Platforms for antibiotic discovery. *Nature Reviews Drug discovery*, 12(5):371. (page 103)

Li, J. W.-H. and Vederas, J. C. (2009). Drug discovery and natural products: end of an era or an endless frontier? *Science*, 325(5937):161–165. (page 103)

Li, M. H., Ung, P. M., Zajkowski, J., Garneau-Tsodikova, S., and Sherman, D. H. (2009). Automated genome mining for natural products. *BMC Bioinformatics*, 10(1):185. (page 6)

Lichten, C. A., White, R., Clark, I. B., and Swain, P. S. (2014). Unmixing of fluorescence spectra to resolve quantitative time-series measurements of gene expression in plate readers. *BMC Biotechnology*, 14(1):11. (page 184)

Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (1997). Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*, 23(1-3):3–25. (page 104)

Liu, W., Luo, Z., Wang, Y., Pham, N. T., Tuck, L., Pérez-Pi, I., Liu, L., Shen, Y., French, C., Auer, M., et al. (2018). Rapid pathway prototyping and engineering using *in vitro* and *in vivo* synthetic genome SCRaMbLE-in methods. *Nature Communications*, 9(1):1936. (pages 190, 194, and 195)

Louvion, J.-F., Havaux-Copf, B., and Picard, D. (1993). Fusion of GAL4-VP16 to a steroid-binding domain provides a tool for gratuitous induction of galactose-responsive genes in yeast. *Gene*, 131(1):129–134. (page 85)

Lubliner, S., Regev, I., Lotan-Pompan, M., Edelheit, S., Weinberger, A., and Segal, E. (2015). Core promoter sequence in yeast is a major determinant of expression level. *Genome Research*, 25(7):1008–1017. (page 58)

Lucas-Abellán, C., Pérez-Abril, M., Castillo, J., Serrano, A., Mercader, M., Fortea, M., Gabaldón, J., and Núñez-Delicado, E. (2019). Effect of temperature, pH, β-and hp-β-cds on the solubility and stability of flavanones: Naringenin and hesperetin. *LWT*, 108:233–239. (page 181)

Luttik, M., Vuralhan, Z., Suir, E., Braus, G., Pronk, J., and Daran, J. (2008). Alleviation of feedback inhibition in *Saccharomyces cerevisiae* aromatic amino acid biosynthesis: quantification of metabolic impact. *Metabolic Engineering*, 10(3-4):141–153. (page 146)

Lyu, X., Ng, K. R., Lee, J. L., Mark, R., and Chen, W. N. (2017). Enhancement of naringenin biosynthesis from tyrosine by metabolic engineering of *Saccharomyces cerevisiae*. *Journal of Agricultural and Food Chemistry*, 65(31):6638–6646. (pages 145, 147, 172, and 193)

Marin, A., Souza, E., Pedrosa, F., Souza, L., Sassaki, G., Baura, V., Yates, M., Wassem, R., and Monteiro, R. (2013). Naringenin degradation by the endophytic diazotroph *Herbaspirillum seropedicae* SmR1. *Microbiology*, 159(1):167–175. (page 148)

McIsaac, R. S., Gibney, P. A., Chandran, S. S., Benjamin, K. R., and Botstein, D. (2014). Synthetic biology tools for programming gene expression without nutritional

perturbations in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 42(6):e48–e48. (pages 57 and 100)

McIsaac, R. S., Oakes, B. L., Wang, X., Dummit, K. A., Botstein, D., and Noyes, M. B. (2013). Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Research*. (pages 86, 87, and 88)

McIsaac, R. S., Silverman, S. J., McClean, M. N., Gibney, P. A., Macinskas, J., Hickman, M. J., Petti, A. A., and Botstein, D. (2011). Fast-acting and nearly gratuitous induction of gene expression and protein depletion in *Saccharomyces cerevisiae*. *Molecular Biology of the Cell*, 22(22):4447–4459. (page 57)

Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischbach, M. A., Weber, T., Takano, E., and Breitling, R. (2011). antismash: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Research*, 39(suppl_2):W339–W346. (page 6)

Mendoza-Ochoa, G. I., Barrass, J. D., Terlouw, B. R., Maudlin, I. E., de Lucas, S., Sani, E., Aslanzadeh, V., Reid, J. A., and Beggs, J. D. (2019). A fast and tuneable auxin-inducible degron for depletion of target proteins in budding yeast. *Yeast*, 36(1):75–81. (page 58)

Migliori, G. B. and Zumla, A. (2017). 148 - Antituberculosis Agents. In Cohen, J., Powderly, W. G., and Opal, S. M., editors, *Infectious Diseases*, pages 1264 – 1276.e2. Elsevier, Fourth edition. (page 107)

Mitchell, L. A., Chuang, J., Agmon, N., Khunsriraksakul, C., Phillips, N. A., Cai, Y., Truong, D. M., Veerakumar, A., Wang, Y., Mayorga, M., et al. (2015). Versatile genetic assembly system (VEGAS) to assemble pathways for expression in *S. cerevisiae*. *Nucleic Acids Research*, 43(13):6620–6630. (page 193)

Mitchell, L. A., Wang, A., Stracquadanio, G., Kuang, Z., Wang, X., Yang, K., Richardson, S., Martin, J. A., Zhao, Y., Walker, R., et al. (2017). Synthesis, debugging, and effects of synthetic chromosome consolidation: synvI and beyond. *Science*, 355(6329):eaaf4831. (page 190)

Mol, J. N., Robbinst, M. P., Dixon, R. A., and Veltkamp, E. (1985). Spontaneous and enzymic rearrangement of naringenin chalcone to flavanone. *Phytochemistry*, 24(10):2267–2269. (page 144)

Mumberg, D., Müller, R., and Funk, M. (1995). Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, 156(1):119–122. (page 60)

Nijveldt, R. J., Van Nood, E., Van Hoorn, D. E., Boelens, P. G., Van Norren, K., and Van Leeuwen, P. A. (2001). Flavonoids: a review of probable mechanisms of action and potential applications. *The American Journal of Clinical Nutrition*, 74(4):418–425. (page 143)

Nussbaum, I., Weindling, E., Jubran, R., Cohen, A., and Bar-Nun, S. (2014). Deteriorated stress response in stationary-phase yeast: Sir2 and Yap1 are essential for Hsf1 activation by heat shock and oxidative stress, respectively. *PLoS ONE*, 9(10):e111505. (page 71)

Organization, W. H. (2015). *Global tuberculosis report 2015*. World Health Organization, 20th edition. (page 105)

Ottoz, D. S. and Rudolf, F. (2018). Constitutive and regulated promoters in yeast: How to design and make use of promoters in *S. cerevisiae*. *Synthetic Biology: Parts, Devices and Applications*, 10:14. (pages 58, 60, and 71)

Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M., Tai, A., Main, A., Eng, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496(7446):528. (page 104)

Partow, S., Siewers, V., Bjørn, S., Nielsen, J., and Maury, J. (2010). Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast*, 27(11):955–964. (page 61)

Pelechano, V., Wei, W., and Steinmetz, L. M. (2013). Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature*, 497(7447):127. (page 58)

Pfleger, B. F., Pitera, D. J., Smolke, C. D., and Keasling, J. D. (2006). Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnology*, 24(8):1027. (page 147)

Pretorius, I. and Boeke, J. (2018). Yeast 2.0—connecting the dots in the construction of the world's first functional synthetic eukaryotic genome. *FEMS Yeast Research*, 18(4):foy032. (page 194)

Quinn, J. Y., Cox III, R. S., Adler, A., Beal, J., Bhatia, S., Cai, Y., Chen, J., Clancy, K., Galdzicki, M., Hillson, N. J., et al. (2015). Sbol visual: a graphical language for genetic designs. *PLoS Biology*, 13(12):e1002310. (page 3)

Rohde, J. R., Trinh, J., and Sadowski, I. (2000). Multiple signals regulate GAL transcription in yeast. *Molecular and Cellular Biology*, 20(11):3880–3886. (page 63)

Ruohonen, L., Penttilä, M., and Keränen, S. (1991). Optimization of Bacillus α-amylase production by *Saccharomyces cerevisiae*. *Yeast*, 7(4):337–346. (page 60)

Sadowski, I., Ma, J., Triezenberg, S., and Ptashne, M. (1988). GAL4-VP16 is an unusually potent transcriptional activator. *Nature*, 335(6190):563–564. (page 85)

Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., and Segal, E. (2012). Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nature Biotechnology*, 30(6):521. (page 69)

Shen, Y., Wang, Y., Chen, T., Gao, F., Gong, J., Abramczyk, D., Walker, R., Zhao, H., Chen, S., Liu, W., et al. (2017). Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science*, 355(6329):eaaf4791. (page 190)

Shinoda, H., Shannon, M., and Nagai, T. (2018). Fluorescent proteins for investigating biological events in acidic environments. *International Journal of Molecular Sciences*, 19(6):1548. (page 180)

Siedler, S., Khatri, N. K., Zsohar, A., Kjærbølling, I., Vogt, M., Hammar, P., Nielsen, C. F., Marienhagen, J., Sommer, M. O., and Joensson, H. N. (2017). Development of a bacterial biosensor for rapid screening of yeast *p*-coumaric acid production. *ACS Synthetic Biology*, 6(10):1860–1869. (page 148)

Siedler, S., Stahlhut, S. G., Malla, S., Maury, J., and Neves, A. R. (2014). Novel biosensors based on flavonoid-responsive transcriptional regulators introduced into *Escherichia coli*. *Metabolic Engineering*, 21:2–8. (pages 148 and 179)

Skjoedt, M. L., Snoek, T., Kildegaard, K. R., Arsovska, D., Eichenberger, M., Goedecke, T. J., Rajkumar, A. S., Zhang, J., Kristensen, M., Lehka, B. J., et al. (2016). Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. *Nature Chemical Biology*, 12(11):951. (pages 147 and 178)

Stanley, R. E., Blaha, G., Grodzicki, R. L., Strickler, M. D., and Steitz, T. A. (2010). The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70s ribosome. *Nature Structural & Molecular Biology*, 17(3):289. (page 107)

Stone, G. and Sadowski, I. (1993). GAL4 is regulated by a glucose-responsive functional domain. *The EMBO Journal*, 12(4):1375–1385. (page 63)

Struhl, K. (1982). The yeast his3 promoter contains at least two distinct elements. *Proceedings of the National Academy of Sciences*, 79(23):7385–7389. (page 58)

Susek, R. E. and Lindquist, S. (1990). Transcriptional derepression of the *Saccharomyces cerevisiae* HSP26 gene during heat shock. *Molecular and Cellular Biology*, 10(12):6362–6373. (page 71)

Thaker, M. N. and Wright, G. D. (2015). Opportunities for synthetic biology in antibiotics: Expanding glycopeptide chemical diversity. *ACS Synthetic Biology*, 4(3):195–206. (pages 103 and 104)

Thomas, M. G., Chan, Y. A., and Ozanick, S. G. (2003). Deciphering tuberactinomycin biosynthesis: isolation, sequencing, and annotation of the viomycin biosynthetic gene cluster. *Antimicrobial Agents and Chemotherapy*, 47(9):2823–2830. (page 108)

Trabelsi, H., Koch, M., and Faulon, J.-L. (2018). Building a minimal and generalizable model of transcription factor–based biosensors: Showcasing flavonoids. *Biotechnology and Bioengineering*, 115(9):2292–2304. (pages 147, 148, 178, 179, 186, and 195)

Tuerk, C. and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, 249(4968):505–510. (page 65)

Uhlmann, T., Boeing, S., Lehmbacher, M., and Meisterernst, M. (2007). The VP16 activation domain establishes an active mediator lacking CDK8 *in vivo*. *Journal of Biological Chemistry*, 282(4):2163–2173. (page 85)

Voyvodic, P. L., Pandi, A., Koch, M., Conejero, I., Valjent, E., Courtet, P., Renard, E., Faulon, J.-L., and Bonnet, J. (2019). Plug-and-play metabolic transducers expand the chemical detection space of cell-free biosensors. *Nature Communications*, 10(1):1–8. (page 195)

Wang, Y., Halls, C., Zhang, J., Matsuno, M., Zhang, Y., and Yu, O. (2011). Stepwise increase of resveratrol biosynthesis in yeast *Saccharomyces cerevisiae* by metabolic engineering. *Metabolic Engineering*, 13(5):455–463. (page 192)

Wassem, R., Marin, A., Daddaoua, A., Monteiro, R., Chubatsu, L., Ramos, J., Deakin, W., Broughton, W., Pedrosa, F., and Souza, E. (2017). A NodD-like protein activates transcription of genes involved with naringenin degradation in a flavonoid-dependent manner in *Herbaspirillum seropedicae*. *Environmental Microbiology*, 19(3):1030–1040. (page 148)

Watts, K. T., Lee, P. C., and Schmidt-Dannert, C. (2004). Exploring recombinant flavonoid biosynthesis in metabolically engineered *Escherichia coli*. *Chembiochem*, 5(4):500–507. (pages 144, 147, 161, 169, and 178)

Weigand, J. E., Sanchez, M., Gunnesch, E.-B., Zeiher, S., Schroeder, R., and Suess, B. (2008). Screening for engineered neomycin riboswitches that control translation initiation. *RNA*, 14(1):89–97. (pages 66, 90, 91, and 100)

Weisshaar, B. and Jenkins, G. I. (1998). Phenylpropanoid biosynthesis and its regulation. *Current Opinion in Plant Biology*, 1(3):251–257. (page 143)

Winkel-Shirley, B. (1999). Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways. *Physiologia Plantarum*, 107(1):142–149. (page 145)

Wright, G. D. (2010). Q&A: Antibiotic resistance: where does it come from and what can we do about it? *BMC Biology*, 8(1):123. (page 104)

Wu, J., Zhou, T., Du, G., Zhou, J., and Chen, J. (2014). Modular optimization of heterologous pathways for *de novo* synthesis of (2S)-naringenin in *Escherichia coli*. *PLoS ONE*, 9(7):e101492. (page 193)

Xiong, L., Zeng, Y., Tang, R.-Q., Alper, H. S., Bai, F.-W., and Zhao, X.-Q. (2018). Condition-specific promoter activities in *Saccharomyces cerevisiae*. *Microbial Cell Factories*, 17(1):58. (page 71)

Xu, P., Li, L., Zhang, F., Stephanopoulos, G., and Koffas, M. (2014). Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proceedings of the National Academy of Sciences*, 111(31):11299–11304. (pages 145 and 147)

Yan, Y., Kohli, A., and Koffas, M. A. (2005). Biosynthesis of natural flavanones in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology*, 71(9):5610–5613. (page 145)

Yim, G., Wang, W., Thaker, M. N., Tan, S., and Wright, G. D. (2016). How to make a glycopeptide: a synthetic biology approach to expand antibiotic chemical diversity. *ACS Infectious Diseases*, 2(9):642–650. (page 105)

Zalatan, J. G., Lee, M. E., Almeida, R., Gilbert, L. A., Whitehead, E. H., La Russa, M., Tsai, J. C., Weissman, J. S., Dueber, J. E., Qi, L. S., and Lim, W. A. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*, 160(1-2):339–350. (page 57)

Zhang, G., Li, Y., Fang, L., and Pfeifer, B. A. (2015). Tailoring pathway modularity in the biosynthesis of erythromycin analogs heterologously engineered in *E. coli*. *Science Advances*, 1(4):e1500077. (page 105)

Zhang, Z. and Dietrich, F. S. (2005). Mapping of transcription start sites in *Saccharomyces cerevisiae* using 5‘ SAGE. *Nucleic Acids Research*, 33(9):2838–2851. (page 58)