

University of Nottingham
UK | CHINA | MALAYSIA

Evolution of Yeast Diversity During Biomanufacturing Processes

Mingzhi Xu

Thesis submitted to the University of Nottingham (UK)

for the degree of Doctor of Philosophy

School of Life Sciences

2025

ABSTRACT

The industrial use of *Saccharomyces cerevisiae* in biomanufacturing is often limited by a decline in productivity during extended fermentation. This thesis investigates the evolutionary dynamics of yeast populations under extended cultivation and explores strategies to enhance production stability. Using diverse yeast strain libraries expressing model intracellular and secreted heterologous proteins, the impact of production burden and natural strain diversity on long-term production performance was first assessed. The secreted protein expression imposed greater fitness costs and was associated with more rapid productivity loss; however, certain genotypes maintained stable expression, highlighting the potential of selecting robust host backgrounds. Further investigation revealed that productivity decline in high-producing strains was primarily driven by mutations in the 2-micron-based production plasmid, particularly in the *MET17* promoter, the heterologous protein coding sequence and/or *FRT* sites. These mutations reduced expression burden and conferred a growth advantage, leading to their fixation over time. In addition, phenotypic heterogeneity was identified as a critical early-stage factor limiting productivity, as high-producing subpopulations exhibited elevated stress levels that limited their persistence in cultures and facilitated the dominance of non- and low-producing variants. Finally, the potential to exploit phenotypic heterogeneity was demonstrated using a feedback-based counterselection system to enrich phenotypically high-producing cells for glutathione (GSH) production. By engineering GSH-responsive promoter *pGSH1* to regulate the counterselectable marker *GAP1*, high-GSH producers were selectively enriched, resulting in an 18%-50% increase in total cellular GSH content. Overall, this thesis provides mechanistic insights into the causes of production instability in yeast cell factories and presents a novel approach to harness phenotypic diversity for

improved metabolite yields. The findings support future efforts to integrate host strain selection, stress management and population control in the development of more stable and productive yeast-based biomanufacturing platforms.

PUBLICATIONS

Xu, M., Vallieres, C., Finnis, C., Winzer, K. and Avery, S. V. (2024). Exploiting phenotypic heterogeneity to improve production of glutathione by yeast. *Microb Cell Fact,* **23**, 267

ACKNOWLEDGEMENTS

Firstly, I would like to thank my principal supervisor, Professor Simon Avery, for his encouragement, support and guidance throughout this project. I also thank my co-supervisors, Dr Klaus Winzer and Dr Chris Finnis, for their advice and support.

I would like to thank my colleagues in B44 for their support, even though many have since moved on and I am now the only member in the lab. A special thank you goes to our lab manager, Matthew Kokolski, for his technical assistance and for ensuring that everything I needed was ordered.

Finally, I thank my two cats, Rich and Milo, for being part of this journey with me, and I thank myself for the persistence and determination that carried me through this PhD. I also wish to thank all the people who came into my life during this four-year journey. Many are now out of touch, but they remain part of the experience that has shaped this fun journey.

从二十岁到三十岁，这最后四年的读博旅程也算是为我自己在英国十年的学生生涯画上了圆满的句号。感谢在我生命中出现过的一切人事物，塑造了现在的我自己。虽然对于未来的人生规划依然有些模糊，但是对于自己所想要的生活轮廓已经渐渐清晰。

人生海海，尽兴就好。

CONTENTS

ABSTRACT ..... *i*

PUBLICATIONS ..... *iii*

ACKNOWLEDGEMENTS ..... *iv*

CONTENTS ..... *v*

LIST OF FIGURES ..... *xii*

LIST OF TABLES ..... *xvii*

LIST OF ABBREVIATIONS AND SYMBOLS ..... *xviii*

1 GENERAL INTRODUCTION *1*

1.1 The development of industrial biomanufacturing **1**

1.1.1 Pre-modern biomanufacturing 1

1.1.2 The industrialization of microbial metabolites 1

1.1.3 The rise of recombinant protein production 2

1.1.4 Current landscape of biomanufacturing 3

1.1.5 Fermentation strategies in modern biomanufacturing 4

1.2 Microorganisms in biomanufacturing and their importance **7**

1.2.1 Overview of microbial biomanufacturing platforms 7

1.2.2 The importance of microbial factories 8

1.2.3 Yeasts in microbial biomanufacturing 10

1.2.4 Basic biology and genetics of *S. cerevisiae* 14

1.3 Cell population heterogeneity in bioprocesses **23**

1.3.1 Genotypic heterogeneity 24

1.3.2 Phenotypic heterogeneity....27

1.4 Culture progression during microbial manufacturing.... 37

1.5 Summary and thesis aims.... 39

1.5.1 Aims....39

2 GENERAL MATERIALS AND METHODS....41

2.1 General growth media preparation and sterilization.... 41

2.2 Plasmids and primers.... 42

2.3 DNA preparation, amplification and analysis.... 42

2.3.1 Plasmid DNA preparation from *E. coli*....42

2.3.2 Yeast genomic DNA extraction....43

2.3.3 Conventional polymerase chain reaction (PCR)....43

2.3.4 Agarose gel electrophoresis....44

2.4 Molecular cloning.... 44

2.4.1 Primer design and *in silico* assembly....44

2.4.2 Vector and insert preparation....44

2.4.3 Gibson assembly and *E. coli* transformation....45

2.4.4 Colony PCR and Sanger sequencing....45

2.4.5 Yeast transformation....46

2.5 Quantification of mCherry expression.... 47

2.6 Flow cytometry and growth measurement.... 47

2.6.1 Flow cytometry....47

2.6.2 Growth rate monitoring....48

2.7 Statistical analyses.... 49

3 PRODUCTION STABILITY AND STRAIN DIVERSITY IN YEAST-BASED HETEROLOGOUS PROTEIN EXPRESSION 50

3.1 Introduction 50

3.1.1 Aims 52

3.2 Chapter-specific materials and methods 53

3.2.1 Yeast strain libraries and culture conditions 53

3.2.2 Ploidy test 55

3.2.3 Multi-generational production monitoring 55

3.3 Results 57

3.3.1 Screening and selection of strains from the libraries 57

3.3.2 Correlation between initial production level and production decline 65

3.3.3 Natural diversity enables the identification of robust and durable high-producing strains 68

3.4 Discussion 70

3.4.1 Secreted proteins impose a greater production burden 70

3.4.2 Multiple determinants of long-term production stability 71

3.4.3 Harnessing natural diversity for selection and engineering of robust strains 72

3.4.4 Conclusions 73

4 GENETIC DRIVERS OF PRODUCTION INSTABILITY IN HIGH-PRODUCING YEAST 74

4.1 Introduction 74

4.1.1 Aims 77

4.2 Chapter-specific materials and methods 78

4.2.1 Yeast strains and culture conditions 78

4.2.2 Curing of the production plasmid from yeast....78
4.2.3 Plasmid extraction from yeast....79
4.2.4 Whole plasmid sequencing....80
4.2.5 Sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE)....80
4.2.6 *MET17* promoter activity assay....81
4.2.7 Quantification of plasmid copy number by droplet digital PCR (ddPCR)....82
4.2.8 Construction of engineered *S. cerevisiae* BY4741 strain expressing FLP recombinase....83
4.2.9 FLP/*FRT*-mediated recombination efficiency assay....84

4.3 Results....86

4.3.1 Differential recovery patterns emerged among evolved strains following repression of production....86
4.3.2 Selection of candidate colonies for further analysis....89
4.3.3 Strain curing and retransformation showed that production decline was primarily due to plasmid-borne factors....92
4.3.4 Whole plasmid sequencing revealed mutations in the production plasmids....94
4.3.5 A single-nucleotide point mutation within the *MET17* promoter can dramatically reduce promoter activity....105
4.3.6 Plasmid copy number changes in the retransformants carrying single-type plasmids....107
4.3.7 *FRT*/FLP-mediated recombination efficiency decreased with the mutated *FRT* sites....111
4.3.8 Natural variation at the *MMS21* locus shows no significant association with production stability but suggests a trend....114

4.4 Discussion....118

4.4.1 The complexity of production decline during extended cultivation....118
4.4.2 Recurrently selected mutations in 2-micron-plasmid expression systems....121

4.4.3 Multiple interacting factors influencing plasmid copy number 124
4.4.4 Strain background effects on plasmid maintenance 125
4.4.5 Conclusions and future work 126

5 PHENOTYPIC DIVERSITY IN HIGH-PRODUCING YEAST STRAINS 128

5.1 Introduction 128
5.1.1 Aims 131

5.2 Chapter-specific materials and methods 133
5.2.1 Strains and cultivation conditions 133
5.2.2 Fluorescence-activated cell sorting (FACS) 133
5.2.3 Construction of unfolded protein response (UPR) reporter system and UPR measurement 134
5.2.4 Oxidative stress measurement 135

5.3 Results 136
5.3.1 Three distinct subpopulations were observed in strains carrying the initial production plasmids 136
5.3.2 The production phenotypes of the three subpopulations were transient, indicating non-heritable 141
5.3.3 Correlation between UPR activation and production at the single-cell level in strains carrying the initial, high-production plasmid 146
5.3.4 Similar population-level UPR activation across strains, but elevated heterogeneity in high producers 149
5.3.5 Correlation between oxidative stress and production at the single-cell level in strains carrying the initial high-production plasmid 152
5.3.6 Elevated population-level oxidative stress and heterogeneity in high producers 154

5.4 Discussion 157

5.4.1 Phenotypic heterogeneity and its link to production instability 157
5.4.2 Intracellular stress as a driver of phenotypic heterogeneity 159
5.4.3 Conclusions and future work 161

6 EXPLOITING PHENOTYPIC HETEROGENEITY TO IMPROVE PRODUCT YIELD..... 164

6.1 Introduction 164
6.1.1 Aims 167

6.2 Chapter-specific materials and methods 168
6.2.1 Yeast strains and cultivation conditions 168
6.2.2 Quantification of total cellular GSH 168
6.2.3 RNA extraction, cDNA synthesis and Real-Time quantitative PCR (RT-qPCR) .169

6.3 Results 172
6.3.1 Construct design for selecting phenotypically high-GSH cells 172
6.3.2 *URA3* is not a suitable counterselectable marker in the system 174
6.3.3 *GAP1* as an alternative counterselectable marker in the system.....187
6.3.4 Cysteine supplementation enhanced the effect of feedback selection on GSH production 199

6.4 Discussion 201
6.4.1 Leveraging phenotypic heterogeneity for production enhancement 201
6.4.2 Impact of marker choice on feedback selection system performance 201
6.4.3 Practical considerations and potential of the feedback selection system 203
6.4.4 Conclusion 205

7 SUMMARY AND CONCLUDING REMARKS..... 206

7.1 Key themes and discussion 206
7.1.1 The fitness-productivity trade-off 206

7.1.2 Mechanisms of production decline 207
7.1.3 Use of phenotypic heterogeneity 208
7.2 Future work 209
7.2.1 Cell cycle contributions to phenotypic heterogeneity 209
7.2.2 Systems-level understanding of production instability 210
7.2.3 Discovery and engineering of robust host genotypes 211
7.2.4 Application beyond the model system 211
REFERENCES 213

LIST OF FIGURES

Figure 1.1: Schematic representation of the three major fermentation strategies used in modern biomanufacturing. 7

Figure 1.2: Sexual reproduction of yeast. 15

Figure 1.3: Population structure of S. cerevisiae showing five clean lineages. 18

Figure 1.4: Genetic map and functional elements of the yeast 2-micron plasmid. 20

Figure 1.5: Four-lineage advanced multigeneration breeding used to generate diverse S. cerevisiae populations for phenotypic screening. 22

Figure 1.6: An example of phenotypic heterogeneity in a clonal yeast population expressing GFP. 28

Figure 1.7: Distributions of single-cell fluorescence intensity illustrating differences in phenotypic heterogeneity. 36

Figure 1.8: illustration of industrial workflow for an extended microbial biomanufacturing campaign. 38

Figure 3.1: Schematic map of the yeast production plasmid. 54

Figure 3.2: mCherry-tagged product expression in selected S. cerevisiae strains from haploid and diploid libraries producing the AP205-VLP-mCherry fusion protein. 58

Figure 3.3: AP205 VLP-mCherry fusion protein expression over time in selected strains from the haploid and diploid libraries. 61

Figure 3.4: mCherry-tagged product expression in selected S. cerevisiae strains from haploid and diploid libraries producing the SARS-CoV-2-mCherry fusion protein. 63

Figure 3.5: SARS-CoV-2 spike fragment mCherry fusion protein expression over time in selected strains from the haploid and diploid libraries. 65

Figure 3.6: Correlation between initial SARS-CoV-2-mCherry fusion protein production level and the magnitude of production decline after 140 generations. 68

Figure 4.1: Expression of the SARS-CoV-2-mCherry fusion protein from evolved cultures following repression of production and re-induction. 88

Figure 4.2: Expression of the SARS-CoV-2-mCherry fusion protein from individual colonies isolated from low-producing evolved cultures after production for 140 generations. 91

Figure 4.3: SARS-CoV-2 mCherry fusion protein production by selected low-producing colonies from High-2 and High-3 cultures after 140 generations, before and after curing and retransformation with early-stage plasmid. 93

Figure 4.4: SARS-CoV-2-mCherry fusion protein expression from single-type plasmids isolated from low-expressing colonies selected from evolved cultures. 96

Figure 4.5: SDS-PAGE analysis of SARS-CoV-2-mCherry fusion protein secretion in culture supernatants from retransformants. 103

Figure 4.6: Specific growth rates of High-2 and High-3 strains at early and late stages of the production cycle, and retransformants with single-type plasmids derived from the late production cycle. 104

Figure 4.7: Comparison of the wild-type MET17 promoter activity with the mutated version. 106

Figure 4.8: Quantification of production plasmid copy number per genome in High-2 and High-3 retransformants. 110

Figure 4.9: Recombination efficiency of FLP recombinase with wild-type and mutated FRT sites. 113

Figure 4.10: Relationship between MMS21 genotype and long-term production stability across 30 selected haploid S. cerevisiae strains. 116

Figure 5.1: SARS-CoV-2-mCherry fusion protein expression profiles at the single-cell level in High-2 and High-3 retransformants. 139

Figure 5.2: Heterogeneity level of mCherry fluorescence in High-2 and High-3 strains carrying the initial production plasmid or the evolved low-producing plasmids, measured by coefficient of variation. 140

Figure 5.3: SARS-CoV-2-mCherry expression levels following regrowth of sorted subpopulations. 143

Figure 5.4: SARS-CoV-2-mCherry fusion protein expression at single-cell level following cell sorting and regrowth of subpopulations. 145

Figure 5.5: Correlation between UPR activation and production in High-2 and High-3 strains carrying the initial production plasmid at the single-cell level. 149

Figure 5.6: Population-level UPR activation and heterogeneity in High-2 and High-3 strains carrying initial or evolved production plasmids. 151

Figure 5.7: Correlation between oxidative stress and production at the single-cell level in High-2 and High-3 strains carrying the initial production plasmid. 153

Figure 5.8: Population-level oxidative stress and heterogeneity in High-2 and High-3 strains carrying initial or evolved production plasmids. 155

Figure 6.1: GSH biosynthesis pathway in S. cerevisiae and feedback regulation of GSH1 expression. 166

Figure 6.2: Design principle of a feedback-based selection system for enriching phenotypically high-GSH-producing cells. 173

Figure 6.3: Construct design in the heterozygous pGSH1-URA3 strain. 175

Figure 6.4: Effect of 5-FOA on cellular GSH content of the engineered S. cerevisiae containing the heterozygous GSH1 locus with URA3 as the counter-selectable marker. 177

Figure 6.5: Effect of 5-FOA on growth of the engineered S. cerevisiae containing the heterozygous GSH1 locus with URA3 as counterselectable marker. 179

Figure 6.6: PCR analysis of the GSH1 locus in colonies derived from long-term 5-FOA treated cultures. 181

Figure 6.7: Construct designs in the homozygous pGSH1-URA3 strain and the pURA3-URA3 control strain. 183

Figure 6.8: Effect of 5-FOA on cellular GSH content of the homozygous engineered S. cerevisiae with URA3 as the counter-selectable marker. 184

Figure 6.9: Cellular GSH level of the control homozygous pURA3-URA3 strain under 5-FOA selection. 186

Figure 6.10: Ura3-GFP expression in the homozygous pGSH1-URA3 strain. .. 187

Figure 6.11: Construct designs in the homozygous pGSH1-GAP1 strain and the pCYC1-GAP1 control strain. 189

Figure 6.12: Effect of D-HIS on cellular GSH content of the homozygous engineered S. cerevisiae with GAP1 as the counter-selectable marker. 191

Figure 6.13: Effect of D-HIS on growth of the homozygous engineered S. cerevisiae with GAP1 as counterselectable marker. 193

Figure 6.14: Cellular GSH level of the control homozygous pCYC1-URA3 strain under 5-FOA selection. 194

Figure 6.15: Lack of heritability of the high-GSH-producing phenotype after removal of selection pressure. 195

Figure 6.16: Effect of D-HIS on GAP1, GSH1, and GSH2 mRNA levels in engineered pGSH1-GAP1 strain. 199

Figure 6.17: Effect of cysteine supplementation on cellular GSH levels in the engineered pGSH1-GAP1 strain. 200

Figure Appendix D.1: Illustration of the minimal FRT site sequence. 271

LIST OF TABLES

Table 4.1: Nature of mutational changes characterised by sequencing of plasmids extracted from the indicated strains after production for 140 generations (the specific sequence changes are detailed in Table Appendix C.1). 99

Table Appendix A.1: All amino acids, nucleotide bases and reagents used in this thesis. 250

Table Appendix B.1: All relevant plasmids used and made in this thesis. 251

Table Appendix B.2: All relevant oligos used in this thesis. 255

Table Appendix C.1: Specific sequence changes identified by whole plasmid sequencing. 267

LIST OF ABBREVIATIONS AND SYMBOLS

SNPs	Single-Nucleotide Polymorphisms
GFP	Green Fluorescent Protein
CV	Coefficient of Variation
MCB	Master Cell Bank
WCB	Working Cell Bank
YPD	Yeast Peptone Dextrose
LB	Luria-Bertani
BMM	Buffered Minimal Media
BMMD + CSM – Leu – Met	Buffered Minimal Media supplemented with Dextrose and Complete Supplement Mixture Dropout Leucine and Methionine
YNB	Yeast Nitrogen Base
h	Hour(s)
min	Minute(s)
s	Second(s)
PCR	Conventional Polymerase Chain Reaction
OD_{600}	Optical Density at 600 nm
OD_{620}	Optical Density at 620 nm
PBS	Phosphate Buffered Saline
SEM	Standard Error of the Mean
R^2	The coefficient of determination
QTL	Quantitative Trait Locus
FRT	Flippase Recognition Target

SDS-PAGE	Sodium Dodecyl-Sulphate Polyacrylamide Gel Electrophoresis
ORF	Open Reading Frame
wt*pMET17*	Wild-type *MET17* promoter sequence
mu*pMET17*	Mutated *MET17* promoter sequence
EV	Empty Vector
Cbf1	Centromere Binding Factor I
UPR	Unfolded Protein Response
ER	Endoplasmic Reticulum
ROS	Reactive Oxygen Species
FACS	Fluorescence-Activated Cell Sorting
H_2DCF-DA	2',7'-dichlorodihydrofluorescein diacetate
DCF	2',7'-dichlorofluorescein
GSH	Glutathione
DTNB	5,5'-dithio-bis-(2-nitrobenzoic acid)
RT-qPCR	Real-Time quantitative Polymerase Chain Reaction
Ct	Cycle Threshold
5-FOA	5-fluoroorotic acid
D-HIS	D-histidine

1 GENERAL INTRODUCTION

1.1 The development of industrial biomanufacturing

1.1.1 Pre-modern biomanufacturing

The use of biological systems including animal cells, plant cells and microorganisms to produce commercially valuable biomolecules and biomaterials is known as biomanufacturing. This field has evolved through successive waves of technological innovation, from ancient fermentation practices to optimized, engineered cell factories [1; 2]. The earliest biomanufacturing technologies may be traced back several millennia, when human beings started to use spontaneous fermentation by mixed cultures of naturally occurring microorganisms to make food and beverages (e.g. bread, yogurt, beer and wine), although they had no knowledge of the nature of the fermentations [3]. Over centuries, empirical knowledge accumulated, laying the foundation for more controlled and systematic fermentation techniques and industrial-scale processes [2].

1.1.2 The industrialization of microbial metabolites

Modern biomanufacturing emerged in the early 20th century with the development of a controlled acetone-butanol-ethanol (ABE) fermentation using *Clostridium* spp. bacteria. This marked the shift from mixed to purified monoculture microbial fermentation and the move to large-scale anaerobic liquid fermentations. The typical products of this period were primary metabolites, which are directly involved in normal cell growth, development and reproduction (e.g. ethanol, butanol, acetone, amino acids and organic acids) [1]. For example, by the 1920s, Pfizer commercialized citric

acid production via fungal fermentation, replacing chemical critic acid synthesis for reasons of cost, safety and sustainability. Similarly, by the mid-20th century, glutamate production transitioned from hydrolysis of wheat gluten to microbial fermentation [1; 4]. Today, the global amino acids market size is more than 30 billion US dollars, with glutamate and lysine ranking as the first- and second-most produced amino acids. Overall, both primarily manufactured by microbial fermentation, and several others, including threonine, leucine, valine and tryptophan are also predominantly produced via fermentation processes [5; 6].

The discovery of penicillin from *Penicillium* fungi by Alexander Fleming in 1928 initiated a second wave of innovation on biomanufacturing, centred on secondary metabolites [1]. Unlike primary metabolites, these compounds are not essential for microbial growth but often provide competitive or defensive advantages [7]. Secondary metabolites such as antibiotics, pigments, and mycotoxins have found wide applications in medicine, agriculture and the food industry [7; 8]. By the 1940s, the industrial-scale production of penicillin became feasible through mutant strain selection and optimization of fermentation conditions [1]. In the following decades, additional antibiotics, including streptomycin and tetracycline, were discovered from microorganisms and introduced as medicines.

1.1.3 The rise of recombinant protein production

The 1970s ushered in the era of modern biomanufacturing with the advent of recombinant DNA technology. Scientists developed methods to insert foreign genes into microbial hosts, allowing these organisms to produce heterologous proteins, which are proteins not naturally synthesized by the host [1]. This breakthrough

transformed drug development by enabling the large-scale production of recombinant therapeutics such as insulin, growth hormones and monoclonal antibodies and vaccines [9]. Today, recombinant protein production constitutes a significant portion of the global biopharmaceutical market. The total biopharmaceutical market size exceeds 500 billion USD, with recombinant proteins contributing over 200 billion USD [10], highlighting the importance of biomanufacturing in modern healthcare.

1.1.4 Current landscape of biomanufacturing

In recent decades, biomanufacturing has entered a new phase driven through advances in synthetic biology, metabolic engineering and genome editing. These fields enable precise control and redesign of cellular metabolism for efficient biosynthesis of both natural and non-natural products. Techniques such as high-throughput screening and modular pathway design and CRISPR-based editing have significantly improved the flexibility and productivity of microbial production platforms [11-13].

While microorganisms were the first expression systems used for industrial biomanufacturing, advances in genetic and metabolic engineering have enabled the development of new expression systems, such as mammalian and plant cells, which provide host-specific post-translational modifications that can be crucial for therapeutic efficacy. These systems complement microbial platforms by enabling production of glycoproteins and complex biologics (e.g. monoclonal antibodies) [9; 14]. Modern biomanufacturing now encompasses a broad spectrum of products, ranging from fermentation-based primary and secondary metabolites to recombinant proteins, industrial enzymes and bio-based polymers [1]. These advancements have positioned

biomanufacturing as a cornerstone of modern biotechnology. It plays a key role in the growing bioeconomy by enabling renewable, sustainable alternatives to traditional chemical synthesis and by driving innovation across the pharmaceutical, food and materials sectors [15].

1.1.5 Fermentation strategies in modern biomanufacturing

Fermentation is a fundamental technology in modern biomanufacturing, enabling the large-scale production of valuable bio-based products through the cultivation of microorganisms or other cell systems. The efficiency, scalability and cost-effectiveness of a bioprocess are closely related to the fermentation strategy employed. Modern industrial fermentation typically follows one of three major modes: batch, fed-batch or continuous fermentation (**Figure 1.1**). Each approach offers distinct advantages and limitations depending on the product type, host system and process requirements [16].

1.1.5.1 Batch fermentation

Batch fermentation is the simplest and most traditional fermentation strategy. It involves the inoculation of cells into a fixed volume of nutrient-rich medium in a closed system [16]. The process proceeds without further input of nutrients, and no material is removed until the end of the run (**Figure 1.1**). Batch processes are widely used in small-scale and early development of bioproduction due to their ease of operation and low risk of contamination [16]. For instance, glutamate and lysine production by *Corynebacterium glutamicum* was originally developed using batch processes [6]. However, batch fermentation suffers from limitations including low productivity and the need for cleaning, sterilization and refilling between batches. Moreover, changing

conditions in the medium as a batch process progresses including nutrient limitation and waste accumulation can lead to variable product yields and inconsistent physiological states of the cells [16; 17].

1.1.5.2 Fed-batch fermentation

Fed-batch fermentation is a semi-continuous process built on the batch concept to overcome limitations of batch fermentation by allowing the periodic or continuous feeding of fresh nutrients during fermentation (**Figure 1.1**). This helps to avoid nutrient depletion and suppress the accumulation of byproducts. By maintaining cells in a more stable metabolic state, fed-batch systems can sustain higher cell densities and extended productivity compared to batch processes [16; 17]. Currently, fed-batch fermentation is widely used in industrial biomanufacturing; for example, in recombinant insulin production with *Saccharomyces cerevisiae* and glutamate and lysine production with *C. glutamicum* [6; 18]. Fed-batch systems usually follow a single-cycle format, with medium added throughout and with harvest at the end, or a repeated-cycle format, where part of the culture is periodically removed and replaced with fresh medium. These are typically variable volume systems. The main drawback of these systems is that a large portion of the fermenter volume remains unused until later stages. This gradual increase in working volume constrains the duration of the productive phase and reduce the overall volumetric productivity [17].

1.1.5.3 Continuous fermentation

Continuous fermentation is an open system where fresh medium is continuously supplied, while culture broth containing cells and product is removed at the same rate (**Figure 1.1**). This allows microbial cultures to remain in a steady physiological state,

and allows for sustained productivity, higher space-time yields, reduced downtime and lower operational costs compared to batch and fed-batch approaches [17]. Continuous fermentation is well established for the large-scale bioproduction of primary metabolites such as ethanol, citric acid and lactic acid [19-21]. A major challenge in long-term continuous operation is maintaining strain stability over time. Issues such as plasmid loss, accumulation of mutations and emergence of non-producing subpopulations can compromise productivity [22]. Despite these challenges, continuous fermentation is considered a highly desirable strategy for industrial biomanufacturing, offering continuous output and cost efficiency. Although not yet widely adopted for biopharmaceuticals, it holds potential for the production of recombinant proteins and therapeutic biologics, especially as advances in process control, strain engineering and automation improve system stability and product consistency.

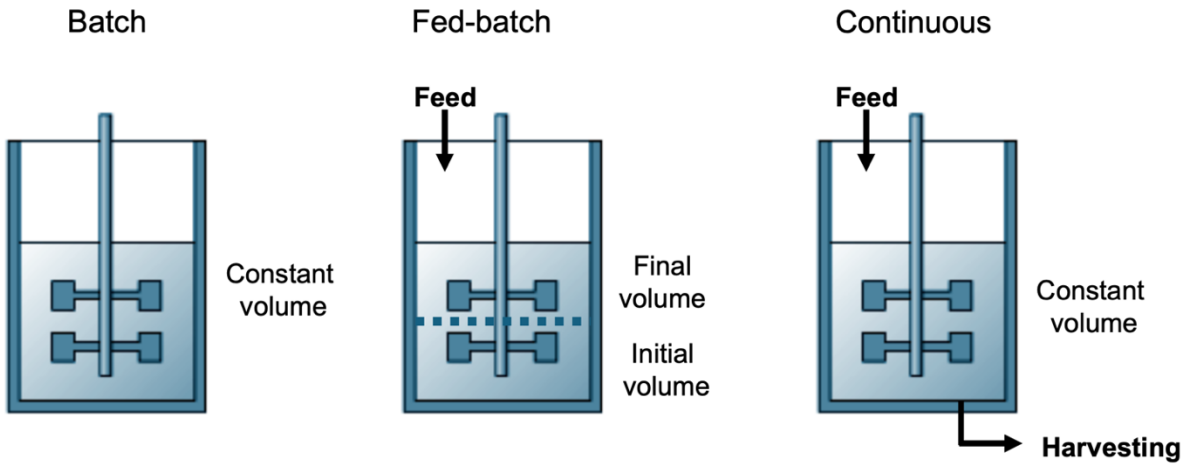


Figure 1.1: Schematic representation of the three major fermentation strategies used in modern biomanufacturing. In batch fermentation, all nutrients are added at the start, and the volume remains constant throughout the process. Fed-batch fermentation involves the addition of fresh nutrients during cultivation, and it can be operated in single-cycle mode (medium added continuously or in pulses until harvest) or cyclic mode (a portion of the culture is periodically removed and replaced with fresh medium). Continuous fermentation maintains a constant volume by continuously adding fresh medium and simultaneously removing culture broth.

1.2 Microorganisms in biomanufacturing and their importance

1.2.1 Overview of microbial biomanufacturing platforms

Microorganisms have played a central role in industrial biomanufacturing for more than a century. Bacteria, filamentous fungi and yeasts are among the most widely used microbial hosts, valued for the ability of some species to grow on low-cost substrates,

adaptability to large-scale fermentation, and capacity to produce a broad range of useful compounds [23]. Bacteria such as *Escherichia coli* and *C. glutamicum* are well-established platforms for producing amino acids, vitamins and recombinant proteins [6; 24; 25]. *E. coli* is particularly favoured for its fast growth, ease of genetic manipulation and high expression levels [25]. Filamentous fungi like *Aspergillus niger* and *Penicillium chrysogenum* are important producers of industrial enzymes and small molecular compounds such as antibiotics and organic acids. These fungi are well known for their excellent ability to secrete large quantities of enzymes and metabolites directly into the culture medium, which simplifies downstream processing and improves overall process efficiency [26]. Yeasts, especially *S. cerevisiae* and *Pichia pastoris*, are also widely employed for manufacturing a diverse range of products, including ethanol, recombinant proteins, vaccines and various high-value metabolites [27; 28]. *S. cerevisiae* remains the most commonly used yeast in biotechnology due to its rapid growth, well-characterized genetics and long-standing safe use in food and beverage industries [9; 28]. Together, bacteria, filamentous fungi and yeasts form the core of microbial platforms that support the global biomanufacturing industry.

1.2.2 The importance of microbial factories

Despite the emergence of alternative expression systems, such as mammalian and plant cells, microbial systems still maintain their overall dominant position in biomanufacturing due to their ease of manipulation, cost-effectiveness, scalability and versatility [9; 29]. Microbes are responsible for the production of a broad range of biologically active compounds across multiple sectors. For example, in the food and feed industries, microbial cell factories are widely used to produce essential amino acids (e.g. glutamate, lysine, threonine), organic acids (e.g. citric acid, lactic acid),

vitamins (e.g. vitamin B2, vitamin C, vitamin K2) and flavour compounds (e.g. vanillin, esters) [6; 24; 30-32]. In the energy sector, microbial biosynthesis provides greener routes to biofuels, offering environmentally friendly alternatives to petroleum-derived products [33]. The pharmaceutical industry also relies heavily on microbial platforms for the production of biologics, including therapeutic proteins, antibiotics and vaccines. Many clinically important molecules, such as penicillin, streptomycin, insulin, interferons and interleukins are manufactured via microbial fermentation using bacteria or fungi [34; 35].

One limitation of microbial hosts generally is a limited capacity for post-translational modifications. This can negatively impact the stability, bioactivity or immunogenicity of certain biologics (e.g. antibodies) [14]. While mammalian expression systems are often favoured for the production of pharmaceutical proteins due to their ability to carry out post-translational modifications similar to those in humans, such as glycosylation, microbes continue to serve as efficient production hosts for many therapeutic proteins, especially for those not requiring elaborate post-translational modifications [9]. Recent advances in synthetic biology and metabolic engineering have further enhanced microbial performance and product quality, enabling broader application of these hosts. For example, human-like *N*-linked glycosylation pathways have been successfully constructed in *E. coli*, making the efficient production of humanized *N*-glycosylated recombinant antibody feasible [36]. Similarly, mammalian *O*-glycosylation pathways have been introduced into *S. cerevisiae* to enable the production of *O*-fucosylated epidermal growth factor domain [37].

1.2.3 Yeasts in microbial biomanufacturing

1.2.3.1 Natural diversity among yeast species and strains

Yeasts are a diverse group of unicellular fungi that have long been utilized in industrial biotechnology. To date, thousands of yeast species have been identified (and each species may comprise hundreds or thousands of recorded strains) but only a relative small number of yeasts have been characterized in detail and employed as cell factories [38]. Among these, *S. cerevisiae* remains the most widely used yeast species in both research and industrial applications. Beyond its traditional role in baking and brewing, *S. cerevisiae* has become a central platform for the production of a wide range of biomolecules. This is largely due to its well-characterised genetic make-up, favourable physiological properties and the availability of a comprehensive set of molecular tools for strain engineering and pathway optimisation [9]. In additional to *S. cerevisiae*, several non-conventional yeasts such as *P. pastoris, Yarrowia lipolytica* and *Hansenula polymorpha*, have been developed as biomanufacturing platforms due to some specific physiological and metabolic advantages [38]. For instance, *P. pastoris* is used widely for the production of recombinant proteins due to its efficient protein secretion system [39]. It is estimated that the yeast species that have been recognized to date only constitute 1% of the extant yeasts on earth [40]. This natural diversity of yeasts, within and among species, can be exploited to screen and select for the most suited industrially relevant traits [41]. For example, a study has described screening and selection of the best-performing *S. cerevisiae* strain for ethanol production, which had an ethanol yield of nearly 85% of theoretical maximum, from a collection of 1313 feral isolates of *S. cerevisiae* [42]. Similarly, another study has shown that incorporation of non-*Saccharomyces* yeasts isolated from wine-related

environments together with *S. cerevisiae* during wine production can improve the quality of wines [43].

1.2.3.2 Laboratory strategies to increase yeast diversity

Despite the enormous biodiversity of wild yeasts, the (combination of) traits giving optimal phenotypes for industrial production sometimes may not be found in nature. Moreover, unlike domesticated yeasts which have potentially lost metabolically costly traits that are not beneficial in industrial settings, wild strains devote cellular resources to survival and reproduction in the natural environment and thus may not perform well in man-made environments [44]. Therefore, a number of laboratory strategies have been developed to increase yeast diversity and select for traits that may enhance biomanufacturing.

One widely used method to increase the yeast diversity is through strain hybridization [45]. Generally, most yeast species can exist in either a haploid or diploid state. Diploid yeast cells (2n) can undergo meiosis to produce haploid spores (n), which can mate with other haploids to regenerate diploids. Crossing haploid strains with different desirable traits generates novel diploid hybrids that can be screened for improved phenotypes. For example, a hybrid diploid *S. cerevisiae* strain TJ14 with enhanced ethanol productivity and high-temperature resistance was created by crossing two parental strains with complementary strengths [46]. For crosses between distantly related species or genera, protoplast fusion is generally used [45]. For example, a hybrid yeast for ethanol production from cheese industry waste was successfully developed by fusing *Kluyveromyces marxianus* with *S. cerevisiae* [47].

Another common method is random mutagenesis using physical or chemical agents, followed by screening. This approach has a long history in strain improvement. For example, random mutagenesis followed by selection was applied to *S. cerevisiae* to isolate mutants with enhanced ethanol tolerance and productivity [48]. Moreover, genetic modification offers a targeted approach to strain improvement. This includes introducing heterologous genes, editing metabolic pathways and manipulating regulatory circuits to optimize yield and efficiency. While genetically modified yeasts face regulatory barriers in food and beverage production, they are widely used in biopharmaceutical manufacturing [45].

Each method described above can generate useful diversity on its own, but combining these approaches often leads to greater improvements. For example, initial hybridization or mutagenesis can broaden diversity, followed by genetic engineering to fine-tune specific metabolic pathways. The ability to utilize both natural and laboratory-derived diversity allows a more flexible strategy for building robust and productive industrial yeast strains.

1.2.3.3 Diploid and haploid yeasts in microbial biotechnology

Both diploid and haploid yeasts are commonly used in microbial biotechnology, depending on the application. Haploid strains are frequently employed in genetic studies, as any mutation introduced is immediately expressed without being masked by a second allele. This makes them particularly useful for gene deletion, pathway optimisation and early-stage strain development [45]. However, haploids tend to be more sensitive to environmental stresses and phenotypic effects of spontaneous mutation, which can compromise their performance under industrial conditions [49]. In

contrast, diploid strains are typically more robust. They tend to show increased tolerance to environmental stress, faster growth rate and improved metabolic stability, which are desirable features for large-scale bioproduction [49; 50]. However, the presence of two gene copies can make strain engineering in diploids more complex, as both alleles must be modified to observe a full phenotype, and recessive traits may be masked [51]. This thesis focuses on the conventional yeast *S. cerevisiae*. Both haploid and diploid *S. cerevisiae* strains were used depending on the experimental objective. Unless otherwise stated, the term 'yeast' hereafter refers specifically to *S. cerevisiae.*

1.2.3.4 Comparison between E. coli and yeast

In comparison to the most used bacterial host *E. coli*, *S. cerevisiae* offers several advantages in terms of biopharmaceutical protein production. *E. coli* was the earliest system used to produce recombinant proteins, and currently still accounts for approximately 25% of total biopharmaceutical protein production [52]. However, its use is often limited due to challenges including protein misfolding, the formation of insoluble inclusion bodies, the presence of endotoxins, poor secretion capacity and the lack of post-translational modification machineries [9; 53; 54]. These issues can compromise product quality and safety, particularly in the context of therapeutic protein manufacturing. In contrast, *S. cerevisiae* is a unicellular eukaryote with more similar sub-cellular structures, post-translational modification, secretion and protein folding pathways to higher eukaryotes. At the same time, it retains several of the practical characteristics that make *E. coli* attractive, including fast growth (albeit not as fast as *E. coli*), ease of cultivation and manipulation, and suitability for high-density

fermentation [9]. Moreover, yeast is endotoxin-free and has been classified as 'generally recognized as safe' [55]. It also has a long history of safe use in industry.

One limitation of yeast is that its glycosylation pattern contains many mannose residues and differs significantly from that in humans, so potentially affecting stability, efficacy and immunogenicity of therapeutic glycoproteins [9]. To address this, extensive glycoengineering efforts have been undertaken to modify the glycosylation pathways of yeast. Examples include the deletion of endogenous mannosyltransferase genes and the introduction of mammalian glycosylation genes, thereby enabling the production of proteins with humanized glycan profiles [37; 56]. In the biopharmaceutical sector, *S. cerevisiae* is now used for the manufacture of several clinically approved protein therapeutics, including human insulin, growth hormone, hepatitis B surface antigen and human papillomavirus virus-like particle [57]. Although yeast currently accounts for only approximately 15% of total biopharmaceutical protein production [52], it remains a highly promising host due to ongoing advances in strain engineering and bioprocess optimisation.

1.2.4 Basic biology and genetics of *S. cerevisiae*

1.2.4.1 Life cycle and mating system

S. cerevisiae is a budding yeast that can grow as either a haploid or diploid. Haploid cells exist as one of two mating types, a or α, and divide asexually by mitotic budding. When haploid cells of opposite mating types encounter each other under suitable nutrient and environmental conditions, then production of mating pheromones is induced. This results in chemotactic growth and fusion of mating partners to form an a/α diploid containing genetic material from both parents. Diploid cells also grow by mitotic budding, but under nutrient limitation they undergo meiosis and sporulation,

producing an ascus that typically contains four haploid spores [58]. Upon return to favourable conditions, spores germinate to form haploid cells, which can either continue vegetative growth or mate to regenerate diploids (**Figure 1.2**).

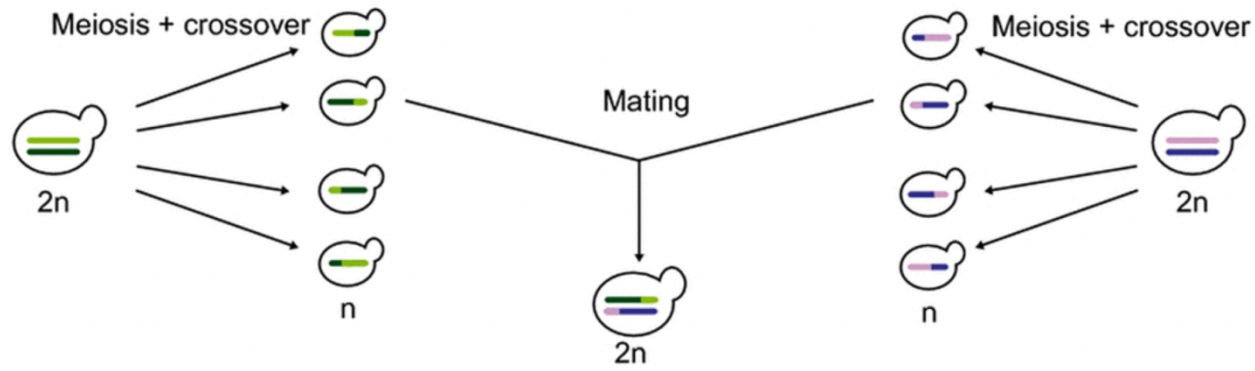


Figure 1.2: Sexual reproduction of yeast. A diploid (2n) yeast cell undergoes meiosis to generate four haploid spores (n). Following germination, haploid cells can mate with cells of the opposite mating type to form a diploid. Figure is taken from Steensels et al., 2014 [45].

A distinct feature of many laboratory and industrial *S.* cerevisiae isolates is homothallism, the ability of a haploid cell lineage to generate both mating types and thereby self-fertilise. Mating identity is determined by the *MAT* locus on chromosome III, while two transcriptionally silent donor loci, HMLα and HMRa, store additional mating-type information. In homothallic strains, the HO endonuclease introduces a site-specific double-strand break at *MAT*, which is repaired by gene conversion using HMLα or HMRa as the template, thereby switching the expressed mating type. Mating-type switching is tightly regulated and occurs under defined cellular conditions rather than being a general response to the environment [59].

Although homothallism can be useful for generating diploids and maintaining sexual reproduction capacity, heterothallic behaviour is often preferred for industrial strain development and deployment. Heterothallism refers to the inability of a haploid cell to switch mating type. In practice, heterothallism is commonly achieved by inactivating the switching machinery, mostly often through loss-of-function of HO [60]. Without functional HO, haploids remain stably a or α, reducing unintended mating and genome reshuffling during cultivation and maintaining engineered genotypes.

1.2.4.2 Genome features and population biology

S. cerevisiae was the first eukaryotic genome to have its genome sequenced. Its nuclear genome consists of 16 chromosomes and is approximately 12.1 Mb in the reference assembly, with around six thousand genes [61]. The reference genome is a useful baseline, but it does not include the extent of variation found across the species. Natural and industrial isolates differ by single-nucleotide polymorphisms (SNPs), indels, gene copy number and larger structural changes. Subtelomeric regions are particularly variable, and changes in chromosome copy number (aneuploidy) are also common, especially under strong selection [62].

This genomic variation is not randomly distributed across the species. Population genomic work has shown that *S. cerevisiae* is genetically structured into lineages, shaped by geography, ecology and domestication history [44; 62; 63]. In a genome-wide survey, five major groups were identified as clean, non-mosaic lineages because they show the same phylogenetic relationship across their entire genomes [63]. These comprise strains from Malaysia, North America, West Africa, sake and related

fermentations (Sake), and a large Wine/European cluster that includes many European and wine-associated isolates (**Figure 1.3**). Many other strains did not fall cleanly within these groups but instead appeared on long branches between the Malaysia and the other clean lineages, consistent with mosaic ancestry arising from historical admixture and recombination, likely facilitated by human-associated dispersal and co-culture in fermentation environments [62]. This is relevant to biotechnology because strain-to-strain differences in traits such as stress tolerance, metabolic capacity and production performance can reflect complex genetic backgrounds rather than variation within a single genetic group. Beyond within-species diversity, interspecific hybridisation is common in industrial environments and has played a major role in brewing. Hybrids often display genome instability, copy-number changes and aneuploid [64], but can also show beneficial trait combinations for fermentation and stress tolerance. For example, larger yeast (*S. pastorianus*), a hybrid derived from *S. cerevisiae* and the cold-tolerant species *S. eubayanus*, supports efficient fermentation at low temperatures [65].

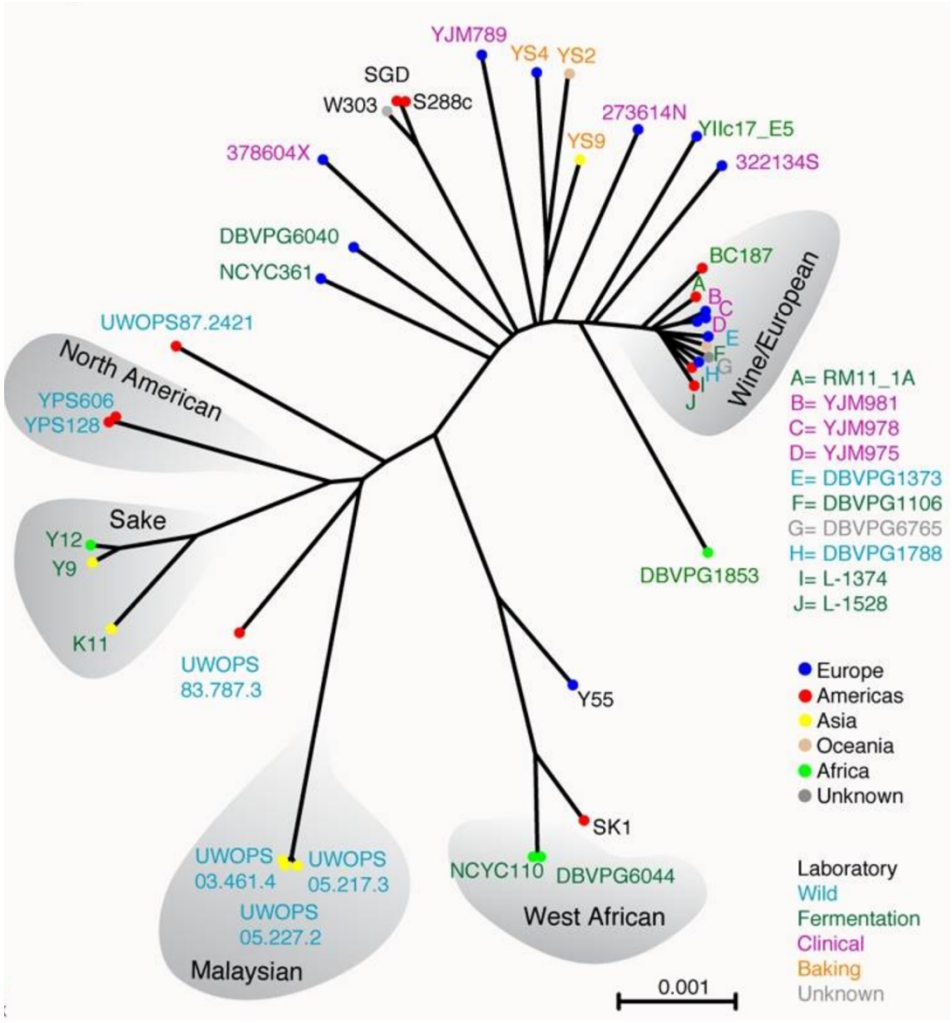


Figure 1.3: Population structure of *S. cerevisiae* showing five clean lineages.

Neighbour-joining tree based on genome-wide sequence variation among representative *S. cerevisiae* strains. Grey shading marks the five clean lineages: Malaysia, North America, West Africa, Sake and Wine/European. Strain names are coloured by isolation source (laboratory, wild, fermentation, clinical, baking or unknown). Dots indicate geographic origin (Europe, Americas, Asia, Oceania, Africa or unknown). Scale bar shows genetic distance. Figure is taken from Liti et al., 2009 [63].

1.2.4.3 The native yeast 2-micron plasmid and its relevance to expression systems

The 2-micron plasmid is a naturally occurring, circular extrachromosomal DNA element found in many *S. cerevisiae* strains. It is maintained at high copy number, typically in the tens of copies per cell and is inherited with low loss rates compared with many engineered episomes [66; 67]. The main genetic features of the native 2-micron plasmid are summarised in **Figure 1.4**. A key feature underpinning high copy number is recombination-based amplification. When plasmid copy number falls below its steady-state level, copy-number recovery is promoted through a feedback mechanism that increases *FLP* expression. Under steady-state conditions, 2-micron gene expression, including *FLP*, is repressed by plasmid-encoded regulators (classically involving Rep1-Rep2, with modulation by Raf1). When copy number drops, this repression is relaxed, leading to increased *FLP* expression and supporting amplification. Additionally, stable inheritance is supported by the *STB* locus together with Rep1 and Rep2, which promotes faithful segregation of plasmid copies to mother and daughter cells and limits plasmid loss across generations [67; 68]. Together, amplification and partitioning enable both high copy number and stable maintenance, making 2-micron-derived vectors useful for relatively high and stable expression in yeast.

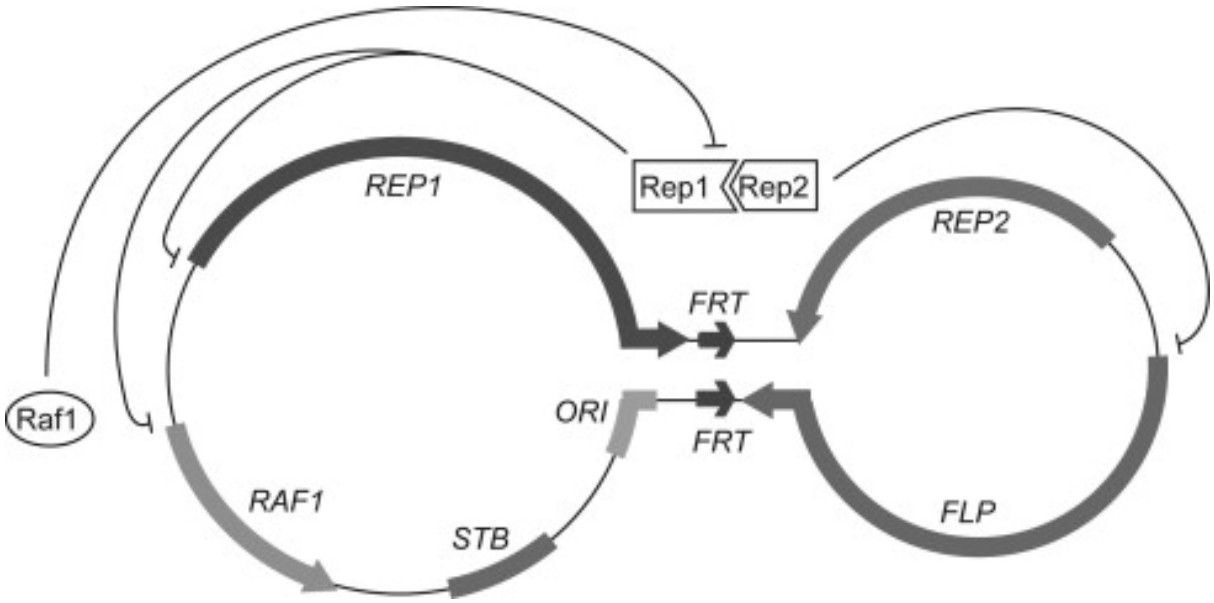


Figure 1.4: Genetic map and functional elements of the yeast 2-micron plasmid.

Diagram of the 2-micron plasmid was drawn as a dumbbell. The central paired segment corresponds to an inverted repeat present in opposite orientations. Major functional sites are indicated, including the replication origin (*ORI*), the segregation locus (*STB*) and the FLP recombination target (*FRT*) sequences. The plasmid encodes four genes: *RAF1*, *REP1*, *REP2* and *FLP*; Arrows indicate transcriptional direction. Rep1 and Rep2 support stable plasmid inheritance by acting at the *STB* partitioning locus and also repress plasmid transcription under steady state conditions. Raf1 counter this repression, permitting expression when required, such as during copy-number recovery. *FLP* encodes a site-specific recombinase that acts at *FRT* sites to catalyse recombination-based amplification and restore plasmid copy number. Figure is taken from Chan et al., 2013 [67].

1.2.4.4 Cutting-edge strain development

As described in Section 1.2.4.2, *S. cerevisiae* contains extensive within-species diversity that can be exploited to improve industrial traits. One practical way to expand

and reshuffle this diversity is advanced multigenerational breeding using multi-parent crosses. In the four-parent advanced intercross framework described in Cubillos et al., 2013, several genetically distinct parental strains are crossed and then cycled through repeated rounds of sporulation and mating [69]. Each generation adds recombination events and breaks up linkage blocks, producing progeny with mosaic genomes assembled from the parental backgrounds.

The strain libraries used in this thesis, provided by Phenotypeca Ltd., were generated using the same general principles [63; 69]. Four haploid parental strains from distinct clean lineages were used (i.e. North American, Sake, West African and European/Wine). The parental strains were first crossed in pairs to form diploids, which were sporulated to generate recombinant haploid progeny of both mating types. Recombinant progenies were then intercrossed to combine all four parental backgrounds, followed by additional rounds of sporulation and mating to increase recombination and generate a panel of advanced intercross lines (**Figure 1.5**). This approach generates a large set of strains with stable, heritable variation distributed across the genome, providing a broad basis for screening and comparing industrially relevant phenotypes under defined conditions. The parental strains were also cured of their native 2-micron plasmid, pre-engineered to enhance recombinant product yield and quality through genetic modifications (e.g., decreased protease expression) and were transformed with high-copy, whole-2-micron-based production plasmids.

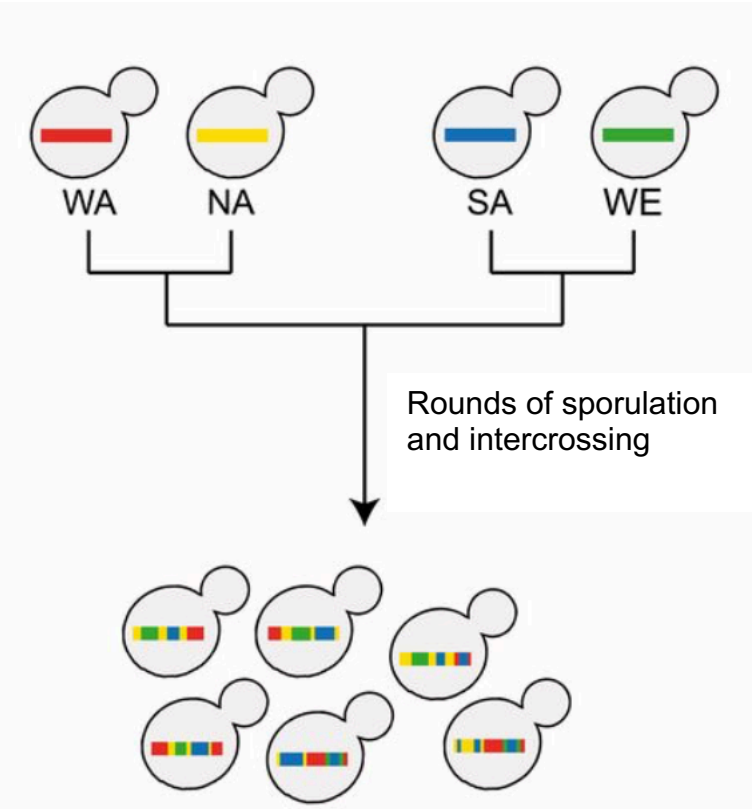


Figure 1.5: Four-lineage advanced multigeneration breeding used to generate diverse *S. cerevisiae* populations for phenotypic screening. Four haploid parental strains representing distinct clean lineages (West African (WA), North American (NA), Sake (SA), and European/Wine (WE)) were crossed in opposite mating-type pairs. The resulting diploids were sporulated, and haploid spores were pooled and subjected to repeated rounds of sporulation and intercrossing to increase recombination and generate diverse recombinant populations. Figure is adapted from Cubillos et al., 2013 [69].

In addition to advanced multigenerational breeding, synthetic genomics offers a way to build yeast strains with fully defined, highly tractable genomes [70]. The Synthetic Yeast Genome Project (Sc2.0) is an international effort to synthesise the *S. cerevisiae* nuclear genome and to assemble redesigned chromosomes with systematic sequence changes. These designs include refactoring of selected sequence features

and insertion of loxPsym sites to enable SCRaMbLE (Synthetic Chromosome Rearrangement and Modificaition by LoxP-mediated Evolution) [71]. When SCRaMbLE is induced, loxPsym recombination generates diverse genome rearrangements, producing large numbers of genotypes from a single synthetic starting strain. Unlike crossing-based approaches that rely on natural meiosis and mating, Sc2.0 provides diversity through engineered chromosome design and inducible genome restructuring.

1.3 Cell population heterogeneity in bioprocesses

Population heterogeneity refers to the variability that exists among individual cells within a clonal cell population [72]. This phenomenon of cell individuality is common in microbial cell cultures and can influence the performance of industrial bioprocesses. In large-scale biomanufacturing, cells are often exposed to suboptimal and heterogeneous environments, including gradients in nutrients, oxygen, pH and the accumulation of toxic byproducts [73-75]. These stresses can amplify cell-to-cell variability by triggering differential stress responses and metabolic states [22; 75]. In addition, random or stress-induced mutations may arise during prolonged cultivation, further contributing to population diversification [76]. Consequently, while some individual cells may maintain high production levels, others may enter low-production or even non-productive states. These imbalances can impact product yields (versus a situation where all cells are performing as well as the best), reducing process efficiency [22]. For example, in a recombinant *S. cerevisiae* chemostat process producing a heterologous insulin product, single-cell sorting showed that 23% of cells had drastically reduced insulin production over around 30 generations [77]. A similar shift has been reported for β-carotene-producing *Yarrowia lipolytica* in continuous

fermentation, where a low-producer subpopulation emerged and β-carotene titer dropped by more than 50% by around 18 growth generations [78]. In some cases, sub-optimally performing subpopulations can come to dominate over time, especially under prolonged or continuous culture conditions [79-81]. Therefore, understanding the nature and causes of this variability is important for designing stable and efficient microbial cell factories. Cell population heterogeneity in bioprocesses can generally be divided into two categories: genotypic heterogeneity, resulting from permanent changes in DNA, and phenotypic (i.e. non-genotypic) heterogeneity, which involves reversible, usually non-heritable differences in cell behaviour.

1.3.1 Genotypic heterogeneity

Genotypic heterogeneity describes heritable differences that arise in microbial populations due to irreversible changes to the DNA. These changes can occur spontaneously and become prominent in response to selective pressures [82]. In microbial bioprocesses, such changes may potentially generate low- and/or non-productive mutants with better fitness, which can come to dominate the population over time, leading to loss of production [79; 83]. Genotypic heterogeneity is especially problematic in long-term or continuous processes and is one of the main challenges for maintaining strain stability [81]. The most common sources of genotypic change include plasmid loss and spontaneous mutation [22].

1.3.1.1 Plasmid loss

Plasmid-based expression systems are commonly used in microbial cell factories because of their ease of construction and potential for high expression levels [84]. However, they are prone to instability, particularly under conditions of high metabolic

burden [85]. Plasmid-free cells often have a growth advantage, especially when the plasmid-driven protein expression imposes a significant load. As a result, those non-producing cells can gradually outcompete productive ones, leading to a reduction in the overall output of the bioprocess. For example, in yeast fermentations, plasmid-free cells have been observed to accumulate during extended batch fermentation, even under auxotrophic selection, with plasmid loss rates reaching up to 56% in 48 h [86]. Interestingly, this plasmid loss was found to be supported by metabolite cross-feeding between the plasmid-carrying cells and plasmid-free cells. When this cross-feeding was disrupted, plasmid retention was improved and recombinant protein expression was increased by 13-fold [86]. A similar phenomenon has been observed in *E. coli*. In engineered *E. coli* producing phycobiliproteins, plasmid loss rates reached 33% within 24 h of induction, despite the use of antibiotic selection [87]. Various strategies have been developed to minimize plasmid loss, including the use of more robust selection systems such as toxin-antitoxin and RNA-based systems in *E. coli*, and defective auxotrophic markers in yeast [86; 87].

1.3.1.2 Spontaneous mutation

Random mutations represent another major source of genotypic heterogeneity in microbial bioprocesses. These genetic changes arise spontaneously during DNA replication or can be induced by certain stress conditions commonly encountered during bioprocesses. For example, oxidative stress generated by high aeration rates has been shown to cause DNA damage, thereby increasing mutagenesis rates [88; 89]. Mutations may include base substitutions, insertions, deletions or larger genomic rearrangements that can affect either the engineered biosynthetic pathway or host regulatory systems [82]. When such mutations confer even a slight fitness advantage,

for example, by reducing the burden of producing a recombinant protein, the affected cells may gradually outcompete their high-producing counterparts. Over time, this evolutionary pressure can shift the population toward low- or non-producing phenotypes and ultimately lead to a progressive decline in productivity during fermentation. In *E. coli* under laboratory conditions, the spontaneous mutation rate has been reported to range from approximately 10^{-10} mutations per base pair per generation for SNPs to around 10^{-5} per gene per generation for mutations mediated by mobile genetic elements [81]. Comparably, *S. cerevisiae* exhibits an overall mutation rate ranging from 3.3×10^{-10} per base pair per generation to around 10^{-5} per gene per generation for recombination-mediated mutation [90; 91]. Although these frequencies appear small, their cumulative impact can become particularly apparent after 50 to 100 generations, which is a common range in industrial settings involving seed culture expansion, semi-continuous cultivation or continuous fermentation [81]. It is important to note that mutation rates can vary across different genomic regions and are influenced by factors such as the host species and gene sequence context [92]. Regions with low GC content, repetitive sequences or strong secondary structures may be more prone to replication errors [93]. Moreover, mutation rates can increase significantly under stress conditions, such as oxygen limitation, nutrient starvation or the accumulation of toxic byproducts [82]. These factors are often present in industrial fermentations. Therefore, engineered strains may be more likely to accumulate genetic changes under industrial settings than in tightly controlled laboratory conditions. Several attempts have been made to alleviate this effect by, for example, reducing the mutation rate or selecting a producer population via metabolic engineering approaches [79; 94; 95].

1.3.2 Phenotypic heterogeneity

While genotypic changes are typically stable and heritable, cell populations can also exhibit substantial non-genotypic variation, even when all cells are genetically identical. This form of variability is referred to as phenotypic (or non-genotypic) heterogeneity. It describes the differences in observable traits, such as gene expression, stress resistance and growth rate, among individual cells within a clonal population [96]. Unlike genotypic heterogeneity, which usually arises over a number of generations when starting from a clonal cell population, phenotypic heterogeneity can always be present in a clonal population. These traits are typically non-heritable (or semi-heritable when there's an epigenetic basis) [96]. A clear illustration of phenotypic heterogeneity can be evidenced with a yeast cell population expressing a green fluorescent protein (GFP) reporter, where cells derived from a single colony display a wide distribution of fluorescence intensities despite being genetically identical (**Figure 1.6**). This type of heterogeneity is thought to play an important role in microbial persistence, stress adaptation, and bet-hedging strategies that enhance population survival under fluctuating conditions [22; 97; 98]. In microbial biomanufacturing, such variation can result in subpopulations that differ in productivity or metabolic activity and thus may reduce the potential efficiency of the process. The main underlying factors that can contribute to phenotypic heterogeneity are typically drivers of heterogeneous gene expression, including cell cycle progress, cell ageing, epigenetic gene regulation, stochastic noise in gene expression and/or cellular components, or microenvironmental variability, as discussed in the following sections.

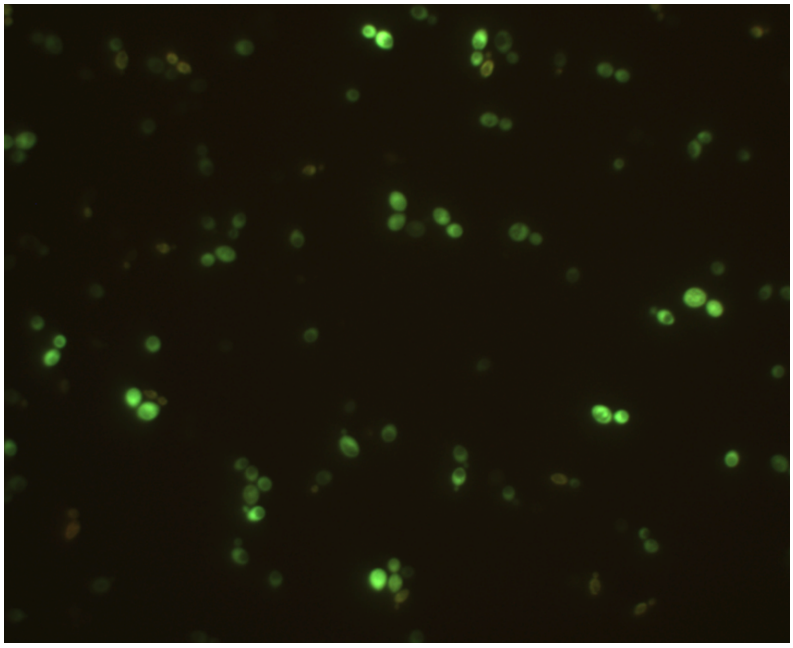


Figure 1.6: An example of phenotypic heterogeneity in a clonal yeast population expressing GFP. Fluorescence microscopy image of genetically identical yeast cells expressing GFP. Despite being derived from a single colony, individual cells display a broad range of fluorescence intensities, illustrating cell-to-cell variation in gene expression. Figure is created by the author.

1.3.2.1 Cellular noise

One of the most universal sources of phenotypic heterogeneity in microbial populations is stochastic fluctuations in gene expression and cellular components, which is known as cellular noise. Even under uniform environmental conditions, individual cells can differ stochastically in their levels of ATP, cofactors, transcription, translation and growth rate [99-103]. These fluctuations can influence systems-level behaviour by propagating across pathways [104]. In the context of microbial

biomanufacturing, such variability can lead to the emergence of subpopulations that differ in productivity, stress resistance, or substrate uptake.

Several biological mechanisms are known to contribute to cellular noise during microbial cultivation. One key source is segregation variability during cell division. When cellular components such as plasmids, transcriptional regulators, transporters or metabolic enzymes are present in low copy numbers, they may be distributed unevenly between daughter cells during division. This unequal partitioning can create distinct phenotypic outcomes immediately after division [22]. For example, the multidrug efflux pump AcrAB-TolC in *E. coli* has been shown to be unequally partitioned at cell division, resulting in persistent variability in antibiotic resistance across a clonal population [105]. Such segregation-based asymmetry can lead to long-lived differences in phenotype among genetically identical cells. Typically, yeast cells divide asymmetrically by budding, producing a larger mother cell and a smaller daughter cell. During this process, aging-related factors such as damaged proteins and defective organelles tend to accumulate in the mother cells, while the newly budded daughter cell inherits a ‘cleaner’ cytoplasmic environment [72]. This asymmetry can lead to physiological differences between the two cell types: mother cells often show slower metabolism, reduced stress tolerance and shorter replicative lifespan, whereas daughter cells, although smaller and slower to initiate growth, generally exhibit greater biosynthetic capacity and are more competent for sustained growth and heterologous protein production [72; 106].

Another major contributor to cellular noise is stochastic gene expression. It arises from fluctuations in transcription and translation that can produce measurable differences

in protein levels among genetically identical cells [102; 107]. This variability is particularly evident in pathways involving low-abundance molecules such as mRNAs or proteins, where even small differences in molecular counts between cells can lead to large differences in phenotypes [96]. Such fluctuations reflect the underlying randomness in the timing and frequency of molecular events within each cell. In this context, two major sources of noise can be distinguished: intrinsic noise, which arises from the probabilistic nature of transcription and translation events at a single gene, and extrinsic noise, which results from cell-to-cell differences in shared regulatory or physiological factors [108]. Studies have shown that intrinsic noise tends to decrease with increasing protein abundance until a minimal threshold is reached [102; 107]. Consequently, essential or regulatory proteins are often produced through strategies that favour high transcription rates and low translation rates, which distribute protein synthesis across many mRNAs and thereby buffer stochastic fluctuations [96]. In contrast, proteins involved in non-essential pathways or secondary metabolism are less tightly regulated and therefore more strongly influenced by extrinsic factors such as promoter activity, stress responses or metabolic state. As a result, these proteins can still display considerable variability, particularly under conditions of strong induction [102]. In microbial production systems, especially when engineering multi-enzyme pathways, gene expression noise can play a critical role by generating imbalances in pathway flux and reducing overall productivity [109-111]. Even if enzymes are expressed at high levels, the activities of different pathway components can still vary between cells. This is particularly true when the genes are located on separate plasmids or expressed using distinct promoters, leading to a lack of coordination in their expression. Experimental and modelling studies have shown that competition for shared resources like ribosomes or RNA polymerases can weaken the

correlation between the expression of different pathway genes in individual cells [112]. The ratio of pathway enzymes can therefore vary unpredictably from cell to cell, which may disrupt pathway efficiency.

Fluctuations in intracellular metabolites and cofactors can also contribute to variability in production. Many biosynthetic pathways depend on cofactors, such as NADH and ATP, but their concentrations can vary significantly between individual cells, even within an isogenic population. For example, studies in *E. coli* have shown substantial variation in intracellular ATP levels among single cells, which can impact synthesis that is ATP-dependent, such as that of lipids [113]. Such differences in energy availability may lead to inconsistent biosynthetic pathway performance across the population, reducing overall yield.

1.3.2.2 Epigenetic gene regulation

Another important source of phenotypic heterogeneity is epigenetic regulation, which refers to heritable but reversible changes in gene expression that do not involve changes to the DNA sequence [96]. In bacteria, such as *E. coli,* the most common type of epigenetic modification is DNA methylation [114]. At the single-cell level, methylation patterns can differ, especially at hypervariable regions of the genome [115; 116]. When these methylation sites occur in or near promoters or regulatory sequences, they can affect the binding of transcription factors and RNA polymerase [114; 117]. As a result, expression of some genes may vary between individual cells in an epigenetic manner, contributing to phenotypic variability. In yeast, which lacks DNA methylation, epigenetic regulation is primarily controlled through histone modifications and nucleosome positioning [118]. These chromatin-based mechanisms

influence the accessibility of transcriptional machinery to DNA. Even within genetically uniform cell populations, studies have shown that accessible chromatin regions can differ between individual cells, largely due to differences in histone modification patterns and nucleosome positioning [119]. For example, the expression of *FLO* genes in yeast, which are responsible for cell adhesion and flocculation, varies widely among cells in a population [120]. This variability is regulated by histone deacetylases, which remove acetyl groups from core histones. This removal restores the positive charge on histone tails, promoting tighter chromatin packing and limiting access of the transcriptional machinery, thereby repressing gene expression [121]. As a result, some cells express certain *FLO* genes and exhibit a particular adhesive behaviour that is not shared by other cells, creating potentially diverse cell subpopulations. Such epigenetic mechanisms give rise to semi-stable expression states, allowing genetically uniform cells to behave differently [117; 119]. In industrial settings, this can affect the uniformity and performance of cells in the culture, especially in continuous or large-scale bioprocesses where minor shifts in population structure can have a cumulative effect on output [22; 122; 123].

1.3.2.3 *Cell cycle stage*

Variation in cell cycle stage is another contributor to phenotypic heterogeneity in microbial populations, especially in eukaryotic systems like yeast. Different stages of the cell cycle are associated with changes in cell size, biosynthetic capacity, and gene expression [123]. Although all cells in a clonal culture go through the same cycle phases, they are usually not synchronized [96]. As a result, individual cells may differ in their physiological state at a given time. In eukaryotic systems, such as yeast, the cell cycle is composed of four distinct phases: Gap 1 (G1), Synthesis (S), Gap 2 (G2)

and Mitosis (M) phases [124]. *S. cerevisiae* cells in the G1 and G2 phases, for example, have been shown to be more resistant to oxidative stress, with G1 linked to menadione resistance and G2 associated with hydrogen peroxide resistance [125]. In contrast, cells during the S phase are more sensitive to hydrogen peroxide, likely due to vulnerability to DNA damage during DNA replication [126]. In a non-cell cycle-synchronized culture, cell subpopulations may also exhibit varying expression of production pathway genes [109; 127]. In prokaryotic systems, such as *E. coli*, the cell cycle does not follow such clearly defined stages as in eukaryotes. However, there is still evidence that cell cycle progression contributes to physiological differences between individual cells. For example, a number of genes have been reported to show dynamic expression patterns across the bacterial cell cycle [128-130]. In addition, fluctuations in intracellular ATP levels have been linked to specific stages of cell growth and division, which may impact biosynthetic efficiency by altering the ATP availability for transcription, translation and metabolite synthesis [100]. When biosynthetic pathways or stress responses are induced at the population level, the resulting expression or response may be uneven, depending on where each cell is in its cell cycle.

1.3.2.4 Microenvironmental variation

Even in well-mixed cultures, spatial gradients of environmental factors, such as oxygen, pH, temperature and nutrient concentrations, can lead to formation of heterogeneous microenvironments [22]. These gradients are more pronounced in large-scale fermenters and can lead to distinct microenvironments that favour different physiological states [131; 132]. As a result, individual cells, may experience different local conditions, giving rise to phenotypic variation in an isogenic population. A recent

study demonstrated the impact of such variability using a proteomics-based approach to track metabolic activity in yeast [133]. Yeasts cells were grown in medium containing ^{13}C-labelled glucose and ^{12}C-labelled lysine. Analysis revealed the emergence of metabolically distinct subpopulations within an isogenic yeast cell population. Some cells predominantly produced lysine, while others consumed it, indicating a metabolic division of labour shaped by nutrient gradients. Moreover, subpopulations undergoing fermentative growth secreted ethanol, which served as a carbon source for nearby respiratory cells. This cross-feeding enabled coexistence of metabolically distinct groups and reflected how local environmental differences can drive functional diversity in an isogenic population. Such metabolic compartmentalization may help the population adapt to local fluctuations [134]. However, in industrial settings, this variability can reduce the predictability and consistency of production processes, reducing the overall performance. For example, cells in oxygen-rich zones in a bioreactor may favour respiration and efficient production (depending on the product), while those in oxygen-limited regions may shift to fermentation, leading to the secretion of ethanol and other byproducts. These uneven physiological states could contribute to inconsistent product quality, reduced yields and difficulties in process control. To mitigate these challenges, industrial processes often rely on strategies such as optimized bioreactor design, enhanced mixing and adaptive process control systems that respond dynamically to internal gradients [135; 136].

1.3.2.5 Measuring phenotypic heterogeneity

One of the most widely used statistical measures for quantifying phenotypic heterogeneity is the coefficient of variation (CV) in a phenotypic value between

indivdiual cells. The CV is defined as the ratio of the standard deviation (σ) to the mean (μ) of a measured parameter and is often expressed as a percentage [137]:

$$\mathrm{CV} = (\frac{\sigma}{\mu}) \times 100\%$$

In single-cell studies, CV is commonly used to describe the distribution of traits such as fluorescence intensity (e.g. reporter of single cell gene expression), enzyme activity, cell size and growth rate [102; 138-140]. The CV value scales with the level of variability between cells (**Figure 1.7**).

In microbial bioprocessing, CV provides a useful metric for assessing the extent of cell-to-cell variability in key production traits. For example, flow cytometry can be used to measure fluorescence intensity across a population expressing a reporter protein, and the CV can then be calculated to quantify gene expression noise. CV is most informative when the underlying data distribution is approximately normal and unimodal. In cases where bimodal distribution or discrete subpopulations exist, complementary approaches, such as mixture modelling and Gini index may be needed to capture the full extent of heterogeneity [140; 141].

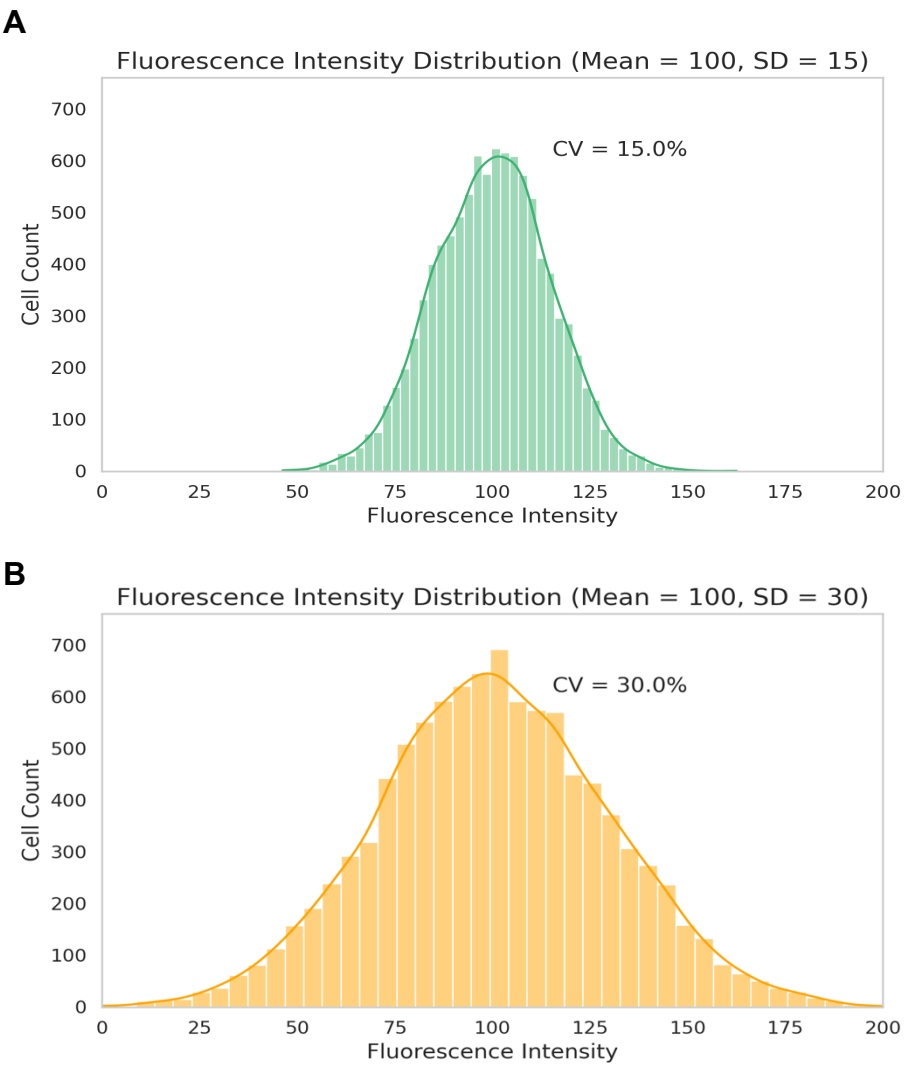


Figure 1.7: Distributions of single-cell fluorescence intensity illustrating differences in phenotypic heterogeneity. (A) Distribution of an isogenic microbial population with a coefficient of variation (CV) of 15%. (B) Distribution of an isogenic microbial population with a coefficient of variation (CV) of 30%. Fluorescence intensity (x-axis) reflects expression levels of a reporter gene; cell count (y-axis) indicates the number of cells with a given expression level. The wider distribution in panel B indicates greater cell-to-cell variability, as represented by a higher CV. NOTE: these are not real data. The figures were generated using random values to illustrate how the CV can be used to describe distribution within an isogenic population. They are intended only as conceptual examples to help explain the principle.

1.4 Culture progression during microbial manufacturing

In large-scale microbial biomanufacturing, particularly for biopharmaceutical production, extended cultivation processes are commonly employed to reduce costs and improve operational efficiency [142]. These extended production campaigns involve the sequential propagation of microbial cells through multiple stages, beginning with a single colony and continuing through seed trains to production vessels (**Figure 1.8**). During this process, yeast cells can undergo more than 100 generations before the final harvest. As illustrated in **Figure 1.8**, an industrial extended production workflow begins with the growth of a single colony on an agar plate, corresponding to approximately 20 generations. This is followed by expansion in a shaking flask, and the culture is used to establish Master Cell Banks (MCB), by which point the cumulative number of generations reaches around 60 generations. From the MCB, a Working Cell Bank (WCB) is prepared via additional passage in liquid culture and subsequent cryopreservation, approximately adding another 10 generations to the cumulative level. These WCB vials are then used to initiate the seed train, which involves a series of scale-up steps to generate sufficient biomass for inoculation of production bioreactors [143]. This typically starts with a shake flask and progresses to a small-scale fermenter, contributing an additional 10 to 20 generations. By the time the culture reaches the final production vessel, the total number of accumulated generations usually exceeds 100.

Although this extended production strategy is cost-effective and favoured in industry, it imposes strong evolutionary pressures on production strains. Over time, even in the presence of deliberate selection, microbial populations can accumulate spontaneous mutations, lose production plasmids and undergo metabolic shifts that reduce

production efficiency [83; 86; 144; 145]. For example, a recent study has shown that in yeast engineered for N-acetylglucosamine production, productivity began to decline after approximately 60-80 generations and was completely lost by 150-200 generations as spontaneous mutations and recombination events resulted in the rise of non-productive mutants with better fitness [79]. These mutants rapidly outcompeted the productive population, ultimately resulting in a complete loss of metabolite output. Similarly, in yeast engineered for the production of a human insulin analogue, long-term cultivation over approximately 80 generations led to a marked decline in productivity, which was linked to metabolic remodelling of the host cells [146]. These findings highlight the vulnerability of engineered production systems to evolutionary drift during extended fermentation campaigns.

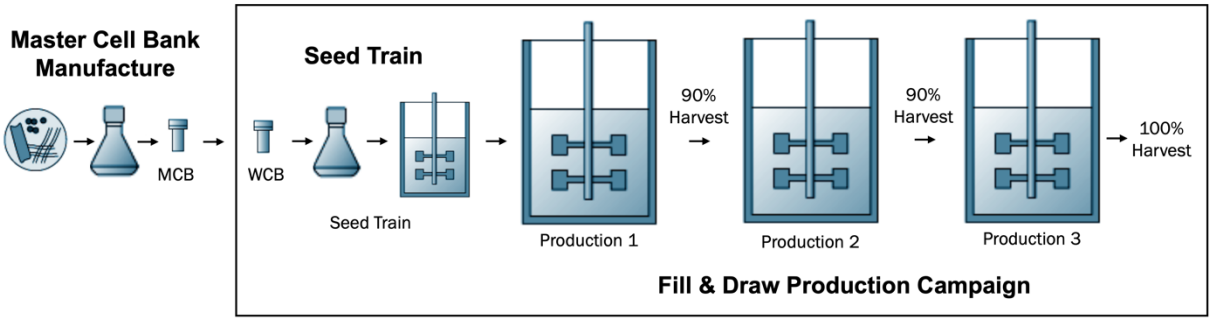


Figure 1.8: illustration of industrial workflow for an extended microbial biomanufacturing campaign. This schematic diagram outlines a typical industrial process for an extended microbial bioproduction campaign, beginning with the isolation of a single colony and progressing through the establishment of a Master Cell Bank (MCB), Working Cell Bank (WCB), seed train expansion and successive fermentation stages.

1.5 Summary and thesis aims

In summary, microbial biomanufacturing has become a central component of the modern bioeconomy, enabling the production of a wide range of high-value products. Among various microbial hosts, *S. cerevisiae* remains a preferred platform due to its eukaryotic nature, genetic tractability, capacity for post-translational modifications and long-standing industrial use. However, challenges remain in maintaining high and stable production outputs, particularly during extended fermentation.

One of the major barriers to consistent bioprocess performance is the emergence of microbial heterogeneity within the production population. Even from a single isogenic colony, individual cells can diverge genetically and/or phenotypically over time. Industrial fermentation processes often involve long cultivation periods, spanning over 100 generations. During this time, cells are exposed to shifting conditions which may accelerate the emergence of low- and/or non-producing variants and ultimately reduce production efficiency. Despite the important commercial implications, few studies have systematically examined how microbial populations evolve during biomanufacturing processes.

1.5.1 Aims

This thesis aims to investigate how heterogeneity in yeast cell populations emerges and evolves during biomanufacturing, specifically as it concerns the use of yeast for the production of high-value proteins, as well as potential solutions to harness the heterogeneity. Specifically, the aims of this thesis are:

1. Test the hypothesis that high-yield production accelerates diversification and selection in yeast culture, and that natural yeast diversity could be exploitable to address this problem.
2. Determine the roles of genotypic heterogeneity and phenotypic heterogeneity in shaping population dynamics during biomanufacturing.
3. Develop and test a strategy to harness phenotypic heterogeneity for yield improvement.

2 GENERAL MATERIALS AND METHODS

2.1 General growth media preparation and sterilization

Media formulations and handling procedures followed established protocols for yeast culture [147; 148]. All liquid media were prepared at the following component concentrations before sterilization: Yeast Peptone Dextrose (YPD) medium was composed of 1% (w/v) yeast extract (Oxoid) and 2% (w/v) bacteriological peptone (Oxoid). Yeast Nitrogen Base (YNB) medium consisted of 0.69% (w/v) yeast nitrogen base without amino acids (Formedium). Buffered minimal (BMM) media consisted of 0.17% (w/v) yeast nitrogen base without amino acids and without ammonium sulphate (Difco), 0.5% (w/v) ammonium sulphate (Sigma), 0.69% (w/v) citric acid (Sigma) and 2.243% (w/v) sodium phosphate, dibasic dihydrate (Thermo Scientific), with pH adjusted to 6.5 using NaOH. Luria-Bertani (LB) medium was prepared with 1% (w/v) tryptone (Oxoid), 0.5% (w/v) yeast extract (Oxoid), 1% sodium chloride (Sigma), with pH adjusted to 7.0 using NaOH. Solid media were prepared by adding agar (Sigma) at the following concentrations to the unsterilized liquid media: 1.6% (w/v) for YPD and YNB media, 2% (w/v) for BMM media and 1.5% (w/v) for LB media.

All media were sterilised by autoclaving at 121 °C, 1.15 bar for 15 min, except for BMM media. Sterilization of liquid BMM media was performed by vacuum filtration using a Corning bottle-top filter unit with a 0.2 µm pore size (Sigma). For the solid form of BMM media, sterilization was conducted by microwave and water bath sterilization method: the medium was first microwaved until the agar was fully dissolved, followed by incubation in a water bath at 80°C for 30 min to ensure sterilization. Unless otherwise specified, D-glucose was added to all yeast media at a final concentration of 2% (w/v)

after sterilization (using a 20% (w/v) glucose stock for YPD and YNB; and a 50% (w/v) glucose stock for BMM). BMM media supplemented with D-glucose are hereafter referred to as BMMD.

Unless otherwise stated, BMMD was supplemented with 0.069% complete supplement mixture dropout leucine and methionine (CSM –Leu –Met) (Formedium), hereafter termed BMMD + CSM –Leu –Met medium. This medium was used for cultivation of strains from the libraries provided by Phenotypeca Ltd.. Laboratory yeast strains were maintained and grown in YPD medium or in YNB medium supplemented as required with amino acids, nucleotide bases and/or reagents from filter-sterilized stock solutions (see **Table Appendix A.1** for details). For cultivation and selection of *E. coli* carrying plasmids of interest, LB medium supplemented with 100 µg/ml carbenicillin was used.

2.2 Plasmids and primers

All plasmids used and constructed in this thesis, as well as all primers, are listed in **Appendix B**.

2.3 DNA preparation, amplification and analysis

2.3.1 Plasmid DNA preparation from *E. coli*

E. coli carrying the desired plasmid was inoculated into 10 mL LB medium containing 100 µg/L carbenicillin and grown at 37 °C, 150 rev. min^{-1} overnight. Plasmid DNA was extracted from *E. coli* using the Nucleospin Plasmid Mini Prep Kit (Macherey-Nagel),

following the manufacturer's instructions. DNA concentration was determined using a Spectrophotometer (DeNovix).

2.3.2 Yeast genomic DNA extraction

Yeast genomic DNA was extracted using the Wizard Genomic DNA Purification Kit (Promega), following the manufacturer's protocol. DNA concentration was determined using a Spectrophotometer (DeNovix).

2.3.3 Conventional polymerase chain reaction (PCR)

Conventional PCR reactions were carried out using a TC-512 thermal cycler (Techne) and Phusion High-Fidelity DNA polymerase (New England Biolabs, NEB). Each 50 µL PCR reaction contained the following components: 10 µL of 5x Phusion HF Buffer, 1 µL of 10 mM dNTPs, 1.5 µL of 10 µM forward primer, 1.5 µL of 10 µM reverse primer, 25-35 ng of plasmid DNA template or 50-100 ng of genomic DNA template, 0.5 µL of Phusion High-Fidelity DNA polymerase, nuclease-free water to a final volume of 50 µL.

The thermal cycling was performed under the following conditions: an initial denaturation at 98 °C for 3 min, followed by 30 to 35 cycles of denaturation at 98 °C for 10 sec (plasmid DNA) or 30 sec (genomic DNA), annealing for 30 sec at the melting temperature (T_m) of the primer with the lower melting point and extension at 72 °C for 30 sec per kilobase of the DNA to be amplified; a final extension step was performed at 72 °C for 10 min.

2.3.4 Agarose gel electrophoresis

DNA samples were separated on 1% (w/v) agarose gels prepared by dissolving agarose in 1x Tris-acetate-EDTA (TAE) buffer (40 mM Tris base, 20 mM acetic acid, 1 mM EDTA) and staining with 1 µg/mL ethidium bromide. Electrophoresis was performed in 1x TAE buffer at 100 V for 30-50 min using a PowerPAC 300 electrophoresis apparatus (Bio-Rad). DNA samples were mixed with 6x Gel Loading Dye, Purple (NEB) to a final concentration of 1x before loading. A 1 kb or 100 bp DNA ladder (NEB) was used as a molecular size marker. Gels were imaged using a Universal Hood Gel Doc System (Bio-Rad).

2.4 Molecular cloning

2.4.1 Primer design and *in silico* assembly

Primer design and *in silico* assembly of constructs were performed using the NEBuilder Assembly Tool (NEB, https://nebuilder.neb.com/#!/). The plasmid backbone pRS315 (ATCC) was used for the construction of all recombinant plasmids generated for DNA manipulation in this study, excluding the production plasmid which was provided by the industrial partner Phenotypeca Ltd. and not derived from pRS315.

2.4.2 Vector and insert preparation

The pRS315 vector was linearised using *Sac*I and *Sal*I restriction enzymes (NEB). The digestion reaction contained: 2 µg of pRS315 plasmid, 2 µL of each enzyme, 5 µL 10x NEB CutSmart Buffer and nuclease-free water to a final volume of 50 µL. The reaction was incubated at 37 °C for 16 h. Desired DNA inserts were amplified by PCR. Both the digested vector and PCR products were separated by agarose gel electrophoresis.

The digested vector was purified using the gel extraction protocol of the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel). PCR-amplified inserts were purified using the PCR clean-up protocol of the same kit, according to the manufacturer's instructions.

2.4.3 Gibson assembly and *E. coli* transformation

Gibson assembly was performed using the NEBuilder HiFi DNA Assembly Cloning Kit (NEB), following the manufacturer's instructions with minor modifications: each assembly reaction was set up using 50 ng of linearised vector and a 5-fold molar excess of each insert, and the reaction mixture was incubated at 50 °C for 1 h to facilitate assembly.

For transformation, *E. coli* XL1-Blue electrocompetent cells were used. A 2 µL aliquot of the assembly product was added to the electrocompetent cells and electroporation was carried out using a MicroPulser Electroporator (BIO-RAD) at 1.8 kV. Following electroporation, cells were recovered with 1 mL LB medium at 37 °C, 150 rev. min^{-1} for 1 h. Cells were then pelleted by centrifugation (1000 x *g*, 5 min), resuspended in 100 µL of sterile water, and the entire volume was spread onto LB agar plates supplemented with 100 µg/L carbenicillin. Plates were incubated at 37 °C overnight.

2.4.4 Colony PCR and Sanger sequencing

Colony PCR was performed by scraping aseptically a small amount of a yeast or *E. coli* colony and suspending it in 40 µL of 20 mM NaOH. The suspension was incubated at 99 °C for 5 min to lyse the cells, followed by brief centrifugation to pellet the cell debris. An aliquot (2 µL) of the resulting supernatant was used as the template for PCR using the standard protocol described in General Materials and Methods 2.3.1.

For *E. coli*, colony PCR was used to screen for correct assembly. Colonies yielding PCR products of the expected size, as confirmed by agarose gel electrophoresis, were further cultivated. Plasmids were then purified and further verified for correct assembly by Sanger sequencing (Eurofins Genomics), following the provider's Sample Preparation Guidelines: 15 µL of sample (plasmid DNA at 50-100 ng/µL or purified PCR products at 5-10 ng/µL) premixed with 2 µL of 10 µM sequencing primer.

For yeast, colony PCR was used to amplify short target sequences (typically < 1000 bp) from genomic DNA or plasmid. Where necessary, PCR products were purified using the NucleoSpin Gel and PCR Clean-up Kit, according to the manufacturer's instructions, and sent for Sanger Sequencing as above.

2.4.5 Yeast transformation

Yeast transformations were performed by using the lithium acetate/single-stranded carrier DNA/polyethylene glycol method [149]: A single yeast colony was inoculated into YPD medium and incubated at 30 °C, 120 rev. min^{-1} (50 mm orbit) overnight. The overnight culture was then diluted to an optical density at 600 nm (OD_{600}) of 0.5 in fresh YPD medium and grown at 30 °C for 3 to 4 h until mid-log phase (OD_{600} of 1.5-2.0) was reached. For each transformation, 5 mL of the mid-log culture was harvested by centrifugation at 1000 x *g* for 2 min at room temperature. The cell pellet was resuspended in 1 mL of Lithium Acetate Tris-EDTA Buffer (100 mM LiAc, 10 mM Tris-HCl pH 8.0, 1 mM EDTA) and incubated at 30 °C for 10 min. Cells were then pelleted again by centrifugation (1000 x *g*, 2 min), and the supernatant was discarded. The resulting cell pellet was resuspended in 300 µL of transformation buffer (100 mM LiAc,

10 mM Tris-HCl pH 8.0, 1 mM EDTA, 40% (w/v) PEG3350). The suspension was then mixed with 52 μL of 2 mg/mL denatured Salmon Sperm DNA (Invitrogen; prepared by boiling at 99 °C for 5 min then immediately cooling on ice) and 1-3 μg of transforming DNA. The transformation mix was incubated at 30 °C for 30 min, followed by a heat shock at 42 °C for 20 min. After heat shock, cells were pelleted by centrifugation (1000 x *g*, 2 min). Where selection required expression of an antibiotic resistance marker, cells were resuspended in 1 mL fresh YPD medium and recovered for 1-2 h at 30 °C, 120 rev. min^{-1} (50 mm orbit). Finally, cells were harvested by centrifugation (1000 x *g*, 2 min), resuspended in 100 μL of sterile deionized water and plated onto appropriately selective agar plates. Plates were incubated at 30 °C for 3-4 days. Successful transformants were verified by PCR and gel electrophoresis using standard procedures as described in General Materials and Methods 2.3.3 and 2.3.4.

2.5 Quantification of mCherry expression

mCherry fluorescence was measured from 500 µL whole culture samples in 48-well microtiter plates at excitation wavelength 540/35 nm and emission wavelength 600/40 nm using a Synergy HTX Multi-Mode Reader (BioTek). OD_{620} was also measured. mCherry fluorescence values were normalized by dividing the fluorescence value by the corresponding OD_{620} value for the culture sample.

2.6 Flow cytometry and growth measurement

2.6.1 Flow cytometry

Yeast cells in exponential growth phase (OD_{600} ~1.5–2.0) were diluted to an OD_{600} 0.5 using sterile phosphate buffered saline (PBS) (137 mM sodium chloride, 2.7 mM

potassium chloride, 11.9 mM phosphate buffer). An aliquot (1 mL) of the diluted sample was washed twice with PBS by centrifugation (1000 x *g*, 2 min) and then resuspended in 1 mL of fresh PBS. Flow cytometry analysis was performed using a CytoFLEX S Flow Cytometer (Beckman Coulter) fitted with a blue (488 nm) laser. Forward scatter (FSC-A) was used to assess cell size. Fluorescent signals were detected in the following channels: yeGFP fluorescence in the B525 channel (525/40 nm emission filter), and mCherry fluorescence in the B585 channel (585/42 nm emission filter). For each sample, 10,000 single cell events were recorded. Fluorescence data were normalized to cell size (FSC-A) and analysed using Kaluza software.

2.6.2 Growth rate monitoring

Yeast growth was monitored using a Synergy HTX Multi-Mode Microplate Reader (BioTek). For standard laboratory strains, 5 µL of overnight culture was transferred into 95 µL of fresh medium in 96-well microtiter plates. For strains derived from Phenotypeca Ltd., 5 µL of overnight culture was transferred into 495 µL of fresh medium in 48-well microtiter plates. Media (BMMD or YNB) were supplemented as necessary for growth and selection. OD_{600} was measured for general laboratory yeast strains (i.e. *S. cerevisiae* BY4741, BY4742 and BY4743), whereas optical density at 620 nm (OD_{620}) was used for strains expressing mCherry to avoid fluorescence interference. Readings were taken every 30 min for 24 to 72 h under continuous linear shaking (1096 cycles min^{-1}, 1 mm orbit) at 30 °C. Specific growth rates were calculated during the exponential growth phase using OD measurements taken over a 3-h time interval.

2.7 Statistical analyses

Data were typically expressed as the mean of at least three biological replicates ± standard error of the mean (SEM). Statistical analyses were performed using using GraphPad Prism 9 software. Comparisons between two groups were assessed using a two-tailed t-test. Comparisons involving more than two groups, or more than one experimental factor, were assessed using one-way or two-way analysis of variance (ANOVA), with an appropriate multiple-comparisons procedure where required. Where ANOVA was used, the test statistic and degrees of freedom are reported as F(df1, df2). For t-tests, t and df are reported. A p-value less than 0.05 ($p<0.05$) was considered statistically significant. Significance is denoted as $p<0.05$(*), $p<0.01$(**), $p<0.001$(***), $p<0.0001$(****).

3 PRODUCTION STABILITY AND STRAIN DIVERSITY IN YEAST-BASED HETEROLOGOUS PROTEIN EXPRESSION

3.1 Introduction

As introduced in Chapter 1, *S. cerevisiae* is a well-established eukaryotic host for industrial manufacturing, offering several advantages for the production of recombinant proteins. Over the past decades, substantial progress has been made in developing synthetic biology tools, high-copy plasmid systems and tuneable expression cassettes to enhance bioproduction in yeast [150-153]. However, maintaining stable production over extended cultivation remains a major challenge that impacts cost-efficiency and process robustness at industrial scales.

A key factor contributing to this instability is how well the expression host performs with any increase in metabolic burden imposed by heterologous protein expression. The expression of heterologous proteins introduces a metabolic burden by diverting cellular resources such as ATP, redox cofactors, amino acids and folding machinery toward non-native pathways [28; 154-156]. This resource allocation can impair cellular homeostasis, trigger stress responses and reduce growth rates [157]. Notably, the magnitude of this burden varies depending on the properties of the expressed protein and metabolic differences between alternative expression hosts. For example, secreted or complex proteins with disulfide bonds and post-translational modifications typically impose higher demands on the secretory pathway, protein folding machinery and intracellular redox systems than simpler, cytosolic proteins [158; 159].

Consequently, strains with naturally higher or engineered chaperone overexpression, e.g. *PDI1*, might perform better than those with low chaperone levels.

In addition to protein-specific burden, the host strain background also plays an important role in shaping expression dynamics. Even when grown under identical conditions, different *S. cerevisiae* strains can vary widely in their production capacity. Such differences arise from natural variation in expression regulation, metabolic capacity and stress tolerance among strains [62; 160]. For example, ethanol yield has been shown to vary by more than 15% among commercial yeast strains [161], and heterologous laccase production can differ up to sevenfold across isolates from laboratory, industrial and wild origins [162]. These findings also suggest that strain-specific factors might either amplify or mitigate the effects of metabolic burden.

Much work has focused on engineering more efficient and resilient yeast cell factories for industrial use [163-165]. However, the yeast strains currently used in industry are often not the best-performing or the most suited, with many being used for historical reasons rather than careful selection and design for a specific application [45; 63]. Recent advances in yeast population genomics have revealed that *S. cerevisiae* harbours enormous natural genetic diversity, which provides a valuable reservoir of traits that can be harnessed to improve strain performance [63]. Combining this natural strain diversity with rational strain engineering offers a new opportunity to select genotypes better suited for industrial applications [45; 166].

3.1.1 Aims

This chapter investigates how heterologous protein burden and strain diversity interact to affect long-term production performance in *S. cerevisiae*. To address this, expression stability was examined across genetically diverse yeast strain libraries using two contrasting heterologous protein candidates: the intracellular AP205 virus-like particle (VLP) and a secreted SARS-CoV-2 spike protein fragment.

Expression trends over time were measured to evaluate how host genotype contributed to production decline. While protein characteristics such as folding or size may also contribute to production stability, these were not directly assessed here. Based on observed differences in production stability, the SARS-CoV-2 spike protein candidate exhibiting a more pronounced decline was selected for further study.

The relationship between initial production yield and the extent of production decline across strains expressing the high-burden protein was examined. This analysis also aimed to assess whether the natural diversity present within *S. cerevisiae* populations could be exploited to identify more robust and stable production hosts for industrial applications.

3.2 Chapter-specific materials and methods

3.2.1 Yeast strain libraries and culture conditions

Haploid (*ura3; LYS2* or *ura3; lys2::URA3*) and diploid (*ura3/ura3; LYS2/LYS2* or *ura3/ura3; LYS2/lys2::URA3* or *ura3/ura3; lys2::URA3/lys2::URA3*) *S. cerevisiae* strain libraries for the production of AP205 VLP and SARS-CoV-2 spike protein fragment were provided by Phenotypeca Ltd. Generation of the libraries are described in Section 1.2.4.4. The libraries had been transformed by Phenotypeca Ltd. with high-copy, whole-2-micron-based plasmids expressing either an intracellular AP205 VLP or a secreted product SARS-CoV-2 spike protein fragment (**Figure 3.1**). Expression of both protein products was under the control of the *MET17* promoter, which allows for regulated expression due to its repression by excess extracellular methionine [167]. Both products carried a C-terminal mCherry tag to facilitate quantitative fluorescence-based monitoring of expression. Hereafter, they are referred to as AP205 VLP-mCherry fusion and SARS-CoV-2-mCherry fusion, respectively. Strains from the libraries were maintained and grown in BMMD + CSM – Leu – Met medium. Yeast cultures (500 µL) were grown in 48-well microtiter plates (Greiner Bio-One) from single colonies at 30 °C, 200 rev. min^{-1} (50 mm orbit).

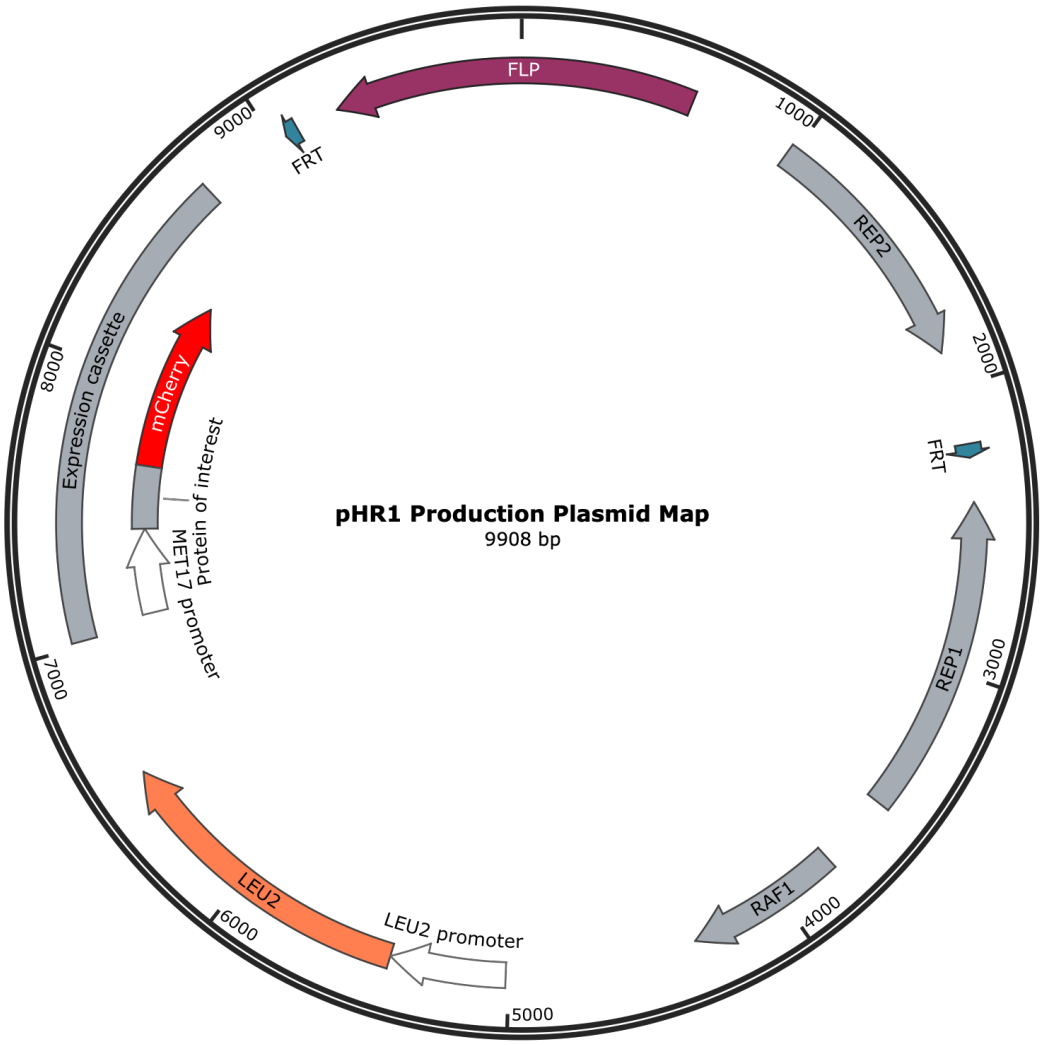


Figure 3.1: Schematic map of the yeast production plasmid. The production plasmid is derived from the endogenous *S. cerevisiae* 2-micron plasmid and contains the whole 2-micron backbone, including the replication and partitioning genes *REP1*, *REP2* and *RAF1*, as well as the *FLP* recombinase gene. The *LEU2* gene serves as a selection marker. The expression cassette was assembled in yeast by in vivo homologous recombination together with the 2-micron sequence during transformation, generating the 2-micron-based production plasmid that drives expression of the recombinant protein of interest under the control of the *MET17* promoter. The protein is C-terminally tagged with mCherry to enable quantitative fluorescence-based monitoring of expression.

3.2.2 Ploidy test

The ploidy of selected strains from the libraries was determined by mating each strain with the test *S. cerevisiae* strains Y55 2369 (*MATα*; *ura2-1; tyr1-1*) and Y55 2370 (*MATa*; *ura2-1; tyr1-1*). For each library strain, three conditions were prepared: (i) the strain grown alone, (ii) co-cultured with Y55 2369, and (iii) co-cultured with Y55 2370. All cultures were grown in 0.5 mL BMMD medium supplemented with 0.1 mg/mL leucine, using 48-well microtiter plates incubated at 30 °C, 200 rev. min^{-1} (50 mm orbit). The plates were sealed with double-layer tape to minimize evaporation. After 3 days of incubation, OD_{620} of the cultures was measured to assess growth. Ploidy was inferred based on the following criteria: the haploid strains only proliferate when co-cultured with a test strain of the opposite mating type, restoring auxotrophy through mating; the diploids either fail to grow under both co-culture conditions or grow independently, regardless of the presence of a test strain. However, this mating-based assay can only confirm the mating-type state at the Mating type (MAT) locus on chromosome III and does not exclude aneuploidy or other karyotypic changes; therefore, definitive ploidy assignment should be confirmed by flow cytometry and/or whole-genome sequencing [168; 169].

3.2.3 Multi-generational production monitoring

For monitoring over many generations, strains were cultivated in 500 µL BMMD + CSM – Leu – Met medium in 48-well microtiter plates at 30 °C, 200 rev. min^{-1} and every 3 days, 2 µL of culture was transferred into a new 48-well microtiter plate containing 498 µL fresh medium. Cultures were measured every 3 days before each transfer for total mCherry fluorescence and OD_{620} as described in **Section 2.5**. Fluorescence values were normalized to OD_{620}. The percentage change in normalised mCherry

fluorescence over the duration of experiment was calculated using the following equation:

$$\left(\frac{Initial\ mCherry\ fluorescence - mCherry\ fluoresnce\ at\ the\ end\ of\ experiment}{Initial\ mCherry\ fluorescence}\right) \times 100\%.$$

3.3 Results

3.3.1 Screening and selection of strains from the libraries

3.3.1.1 *Screening and selection of strains from AP205 VLP* libraries

Previous studies have shown that engineered microbial strains designed for high-level bioproduction often experience reduced cellular fitness due to the metabolic burden associated with bioproduction [80; 170; 171]. Based on this, it was hypothesized in the present study that strains with higher initial production yields may be more prone to selection pressure over time, leading to the emergence and eventual dominance of low-producing or more fit variants. Cells with changes resulting in faster growth due to reduced burden from lower productivity could outgrow other cells in the population over time resulting in a progressive decline in overall productivity. To explore this hypothesis, 72 strains were first screened and selected from both the diploid and haploid libraries with the production plasmid encoding the AP205 VLP-mCherry fusion protein (i.e. 144 strains in total), provided by Phenotypeca Ltd. Both libraries were included to examine whether ploidy influences production stability, as diploid strains are often thought to be more robust and genetically stable than haploids [172; 173]. The strains were isolated by diluting and spreading the library stocks provided by Phenotypeca onto BMMD + CSM – Leu – Met agar plates. Colonies were selected based on visible mCherry fluorescence intensity, with redder colonies presumed to be high producers and paler colonies considered likely low producers. mCherry fluorescence of all 144 strains was then quantitatively measured using a plate reader to confirm their classification as high or low producers (**Figure 3.2**). The haploid library had been prepared by Phenotypeca Ltd. through sporulation of the diploid library, followed by diethyl ether treatment, a method used to selectively eliminate non-sporulated diploid cells [174]. However, as this method does not ensure complete

removal of diploids, ploidy testing was also conducted on selected strains from the haploid library to verify their haploid status.

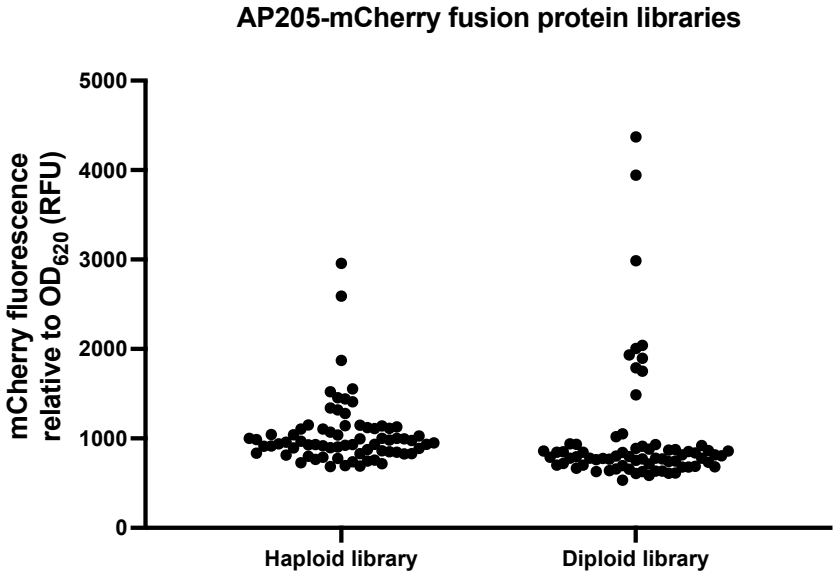
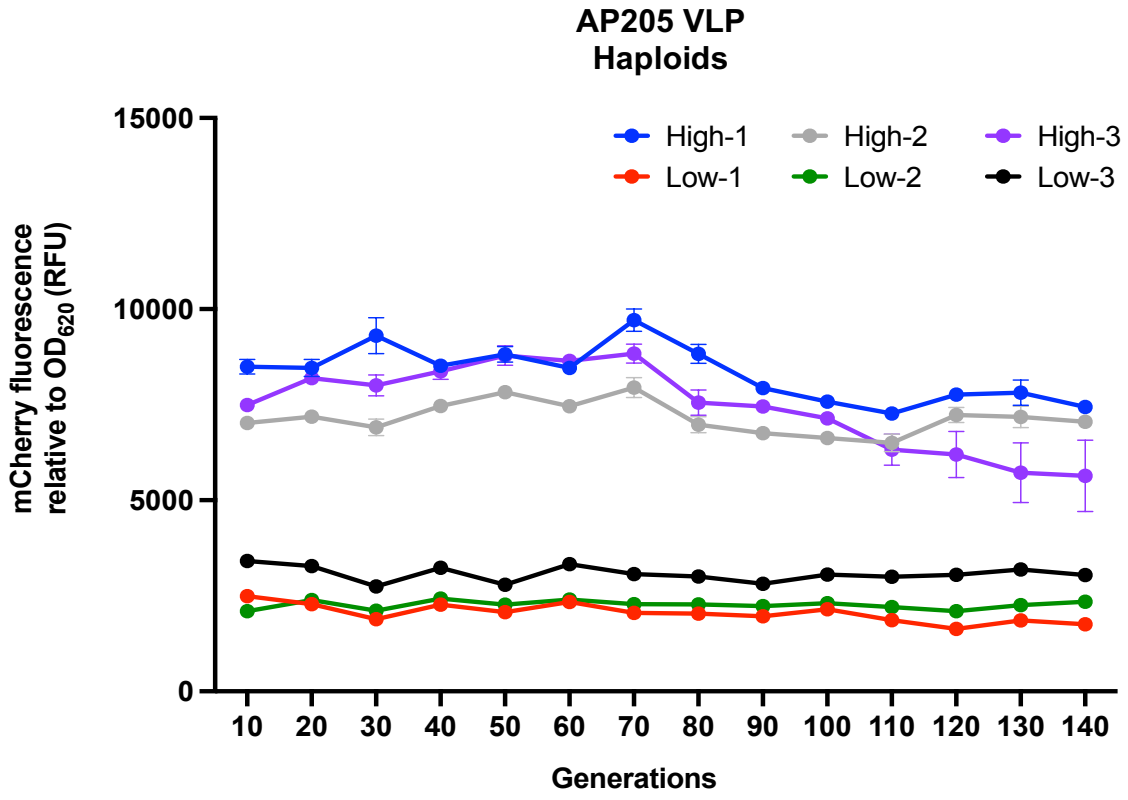


Figure 3.2: mCherry-tagged product expression in selected *S. cerevisiae* strains from haploid and diploid libraries producing the AP205-VLP-mCherry fusion protein. 72 haploid and 72 diploid strains (144 total) were analysed. mCherry fluorescence was measured using a Tecan plate reader (different from the plate reader used for the main results). Each data point represents an individual strain.

To assess long-term production performance, three high-producing strains and three low-producing strains from each library were selected for further extended cultivation and monitoring. For each selected strain, eight biological replicates (i.e. eight clones, where a clone of a strain refers to a clonal colony derived from a single cell of that strain) were cultivated in 48-well microtiter plates. The eight biological replicates were measured to allow more reliable comparison between the different genetic backgrounds and to overcome potential issues from the random timings of specific

changes in a single replicate, masking fundamental genomic differences between the strains. Cultures were subcultured every 3 days into fresh medium to an initial OD_{620} ~0.05, corresponding to approximately 10 generations per passage. mCherry fluorescence was measured before each transfer, enabling the monitoring of AP205 VLP-mCherry production over the course of 140 generations. Expression levels were first recorded after the initial 3-day growth period (i.e. after ~10 generations), as this allowed cells to reach the stationary phase and reflect their early production capacity. At this point, an approximate 4-fold difference in AP205 VLP-mCherry expression was observed between the highest- and lowest-producing strains in both selected haploids and diploids (**Figure 3.3**). Those high producers maintained significantly elevated expression levels compared to those low producers, and no striking decline in mCherry fluorescence was observed over time. These findings indicate that the intracellular expression of the AP205 VLP-mCherry fusion did not impose a sufficient metabolic burden to trigger a rapid evolutionary selection against high production. This reinforces the point that some proteins and the heterologous expression constructs inserted into *S. cerevisiae* to produce them do not present significant issues to long-term stability during biomanufacturing campaigns. Nevertheless, a slight productivity decline was observed for the high producers, suggesting that they still may not be optimal for long-term productivity.

A


AP205 VLP
Haploids
High-1
High-2
High-3
Low-1
Low-2
Low-3
mCherry fluorescence relative to OD620 (RFU)
15000
10000
5000
0
10 20 30 40 50 60 70 80 90 100 110 120 130 140
Generations

B

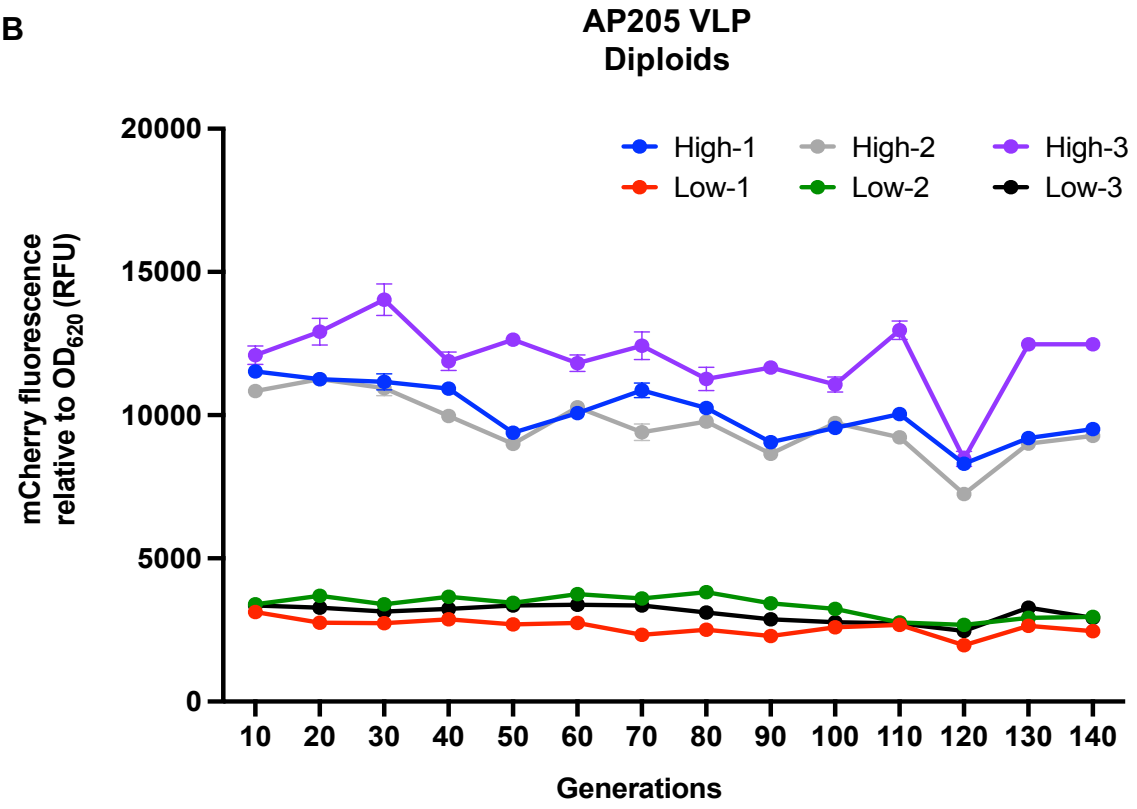

AP205 VLP
Diploids
High-1
High-2
High-3
Low-1
Low-2
Low-3
mCherry fluorescence relative to OD620 (RFU)
20000
15000
10000
5000
0
10 20 30 40 50 60 70 80 90 100 110 120 130 140
Generations

Figure 3.3: AP205 VLP-mCherry fusion protein expression over time in selected strains from the haploid and diploid libraries. mCherry level was measured and normalized to OD_{620} every ~10 generations during sequential subculture into fresh medium every 3 days. Shown are expression profiles over 140 generations for three high-producing strains (High-1, High-2, High-3) and three low-producing strains (Low-1, Low-2, Low-3) from the haploid (A) or diploid (B) library expressing AP205 VLP-mCherry fusion protein. Each data point represents the mean value ± SEM of eight biological replicates (i.e. eight clones derived from a single strain stock).

3.3.1.2 Screening and Selection of strains from SARS-CoV-2 protein libraries

To further explore the relationship between production burden and evolutionary dynamics, a secreted heterologous protein, SARS-CoV-2 spike protein fragment fused to mCherry, was used for subsequent experiments. The spike protein fragment was the SARS-CoV-2 Receptor Binding Motif (RBM) fragment (residues 438-505) with mCherry genetically fused to the C-terminus. This comprises the region of the spike protein that binds to the ACE2 receptor, containing residues forming the disulphide bond Cys480-Cys488 in the full-length spike protein, but lacks N-linked glycosylation motifs. An N-terminal Suc2 invertase secretion signal peptide was included upstream of the SARS-CoV-2-mCherry fusion to target the protein to the secretion pathway and support secretion to the culture supernatant. Deliberate over-secretion of heterologous proteins is known to place additional demands on *S. cerevisiae*, which typically has a low secretory capacity, having primarily evolved to metabolise fruit sugars, while producing ethanol and acetate to inhibit microbial competitors. Consequently, during the high-level secretion of recombinant proteins, the protein folding machinery and

intracellular redox systems are likely to impose a higher metabolic burden than for cytosolic expression alone [158; 159]. In addition, most industrially important therapeutic proteins are secreted by baker's yeast, e.g. insulin and GLP-1 analogues, with relatively few, e.g. HBsAg vaccine VLPs, being produced intracellularly [18; 52; 57].

Using a similar procedure to that of the AP205 VLP, 72 strains were randomly selected from each of the haploid and diploid SARS-CoV-2-mCherry-producing libraries. mCherry fluorescence was quantified in all 144 strains to assess their expression levels (**Figure 3.4**). Ploidy testing was also performed to verify the genetic status of strains classified as haploid. Based on the expression data, three high-producing and three low-producing strains from each library were selected for further extended cultivation and long-term monitoring. At the initial cultivation period (i.e. 10th generation), an approximate 6-fold difference in SARS-CoV-2-mCherry expression was observed between the highest and lowest-producing strains in the selected haploids and diploids (**Figure 3.5**). In contrast to the intracellular AP205 VLP-mCherry producing strains, which maintained relatively steady expression levels across 140 generations (**Figure 3.3**), some high-producing SARS-CoV-2-mCherry strains from both haploid and diploid libraries displayed a marked, progressive decline in expression over 140 generations (**Figure 3.5**). This trend was particularly evident in High-2 and High-3 strains from the haploid library (**Figure 3.5A**) and in High-1 and High-2 strains from the diploid library (**Figure 3.5B**), where expression levels dropped close to those of the low-producing strains. In contrast, the High-1 haploid and the High-3 diploid appeared to have a significantly greater stability over the 140-generation cultivation period and did not decline to the low expression level observed

for the other strains. The low-producing strains from both libraries maintained relatively stable, albeit low-level expression, throughout the cultivation period.

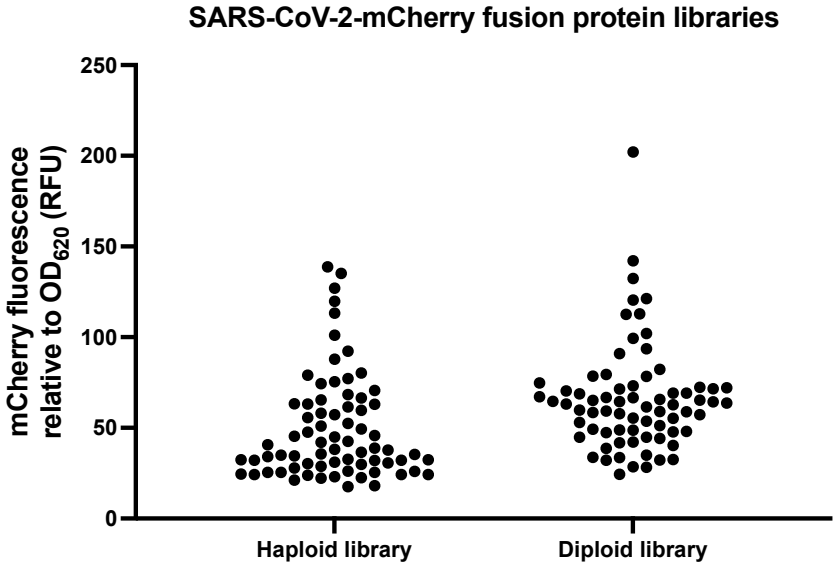
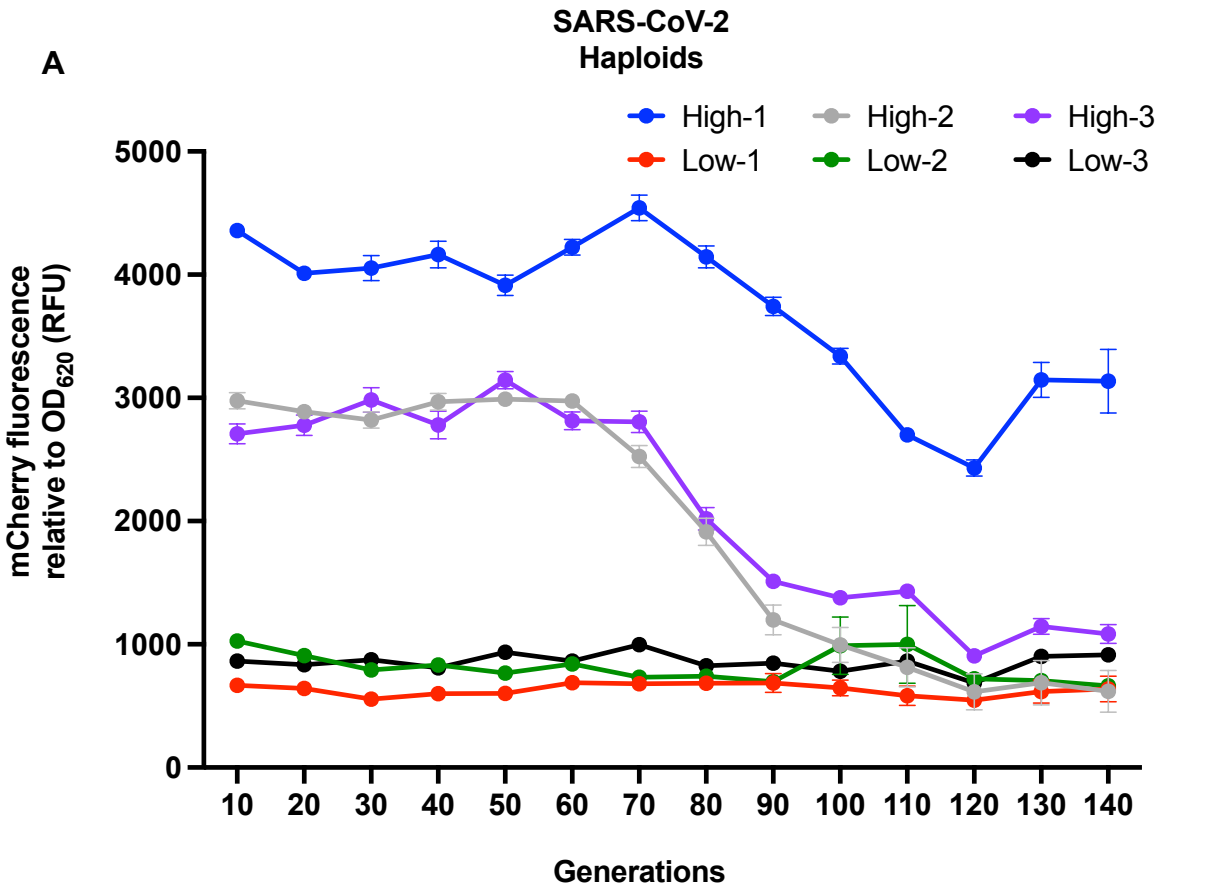


Figure 3.4: mCherry-tagged product expression in selected *S. cerevisiae* strains from haploid and diploid libraries producing the SARS-CoV-2-mCherry fusion protein. 72 haploid and 72 diploid strains (144 total) were analysed. mCherry fluorescence was measured using a Tecan plate reader (different from the plate reader used for the main results). Each data point represents an individual strain.


A
SARS-CoV-2
Haploids
High-1
High-2
High-3
Low-1
Low-2
Low-3
mCherry fluorescence relative to OD620 (RFU)
5000
4000
3000
2000
1000
0
10 20 30 40 50 60 70 80 90 100 110 120 130 140
Generations

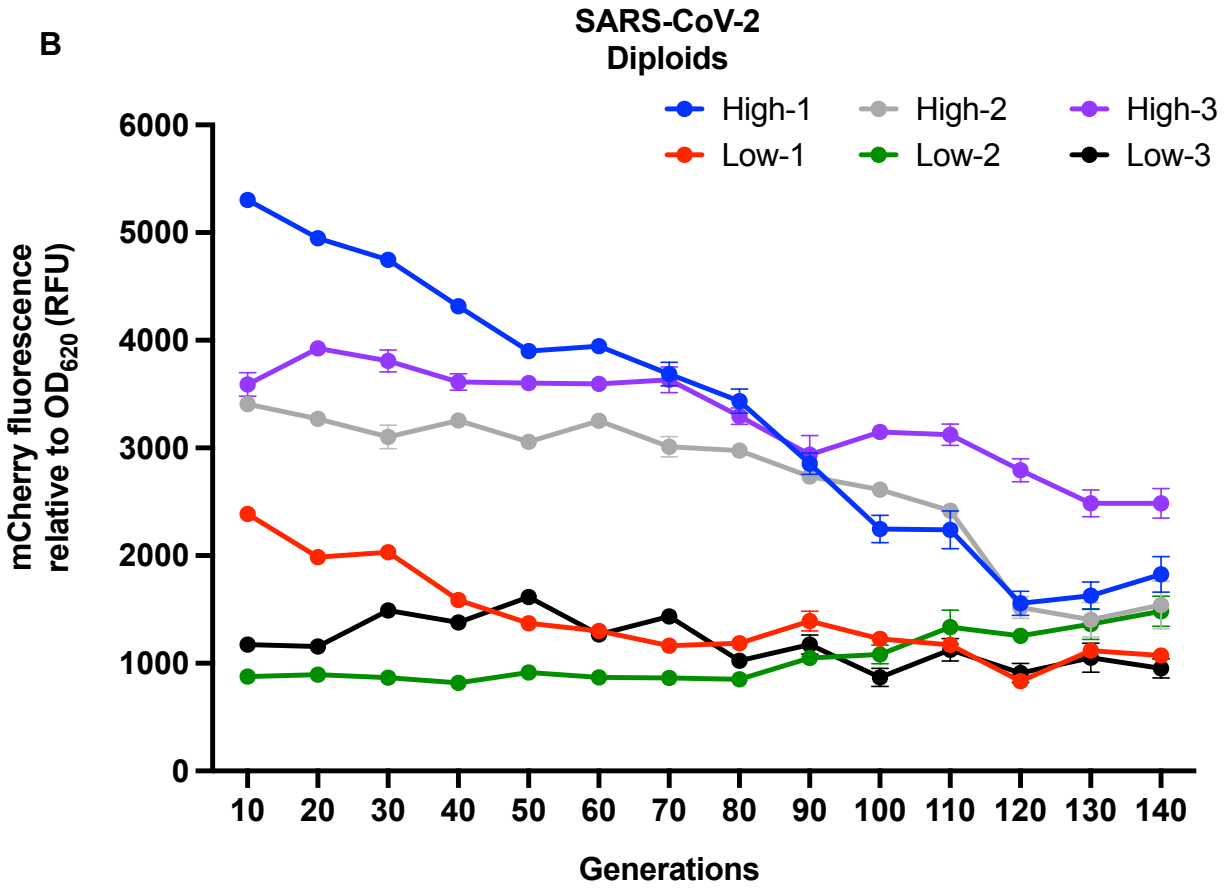

B
SARS-CoV-2
Diploids
High-1
High-2
High-3
Low-1
Low-2
Low-3
mCherry fluorescence relative to OD620 (RFU)
6000
5000
4000
3000
2000
1000
0
10 20 30 40 50 60 70 80 90 100 110 120 130 140
Generations

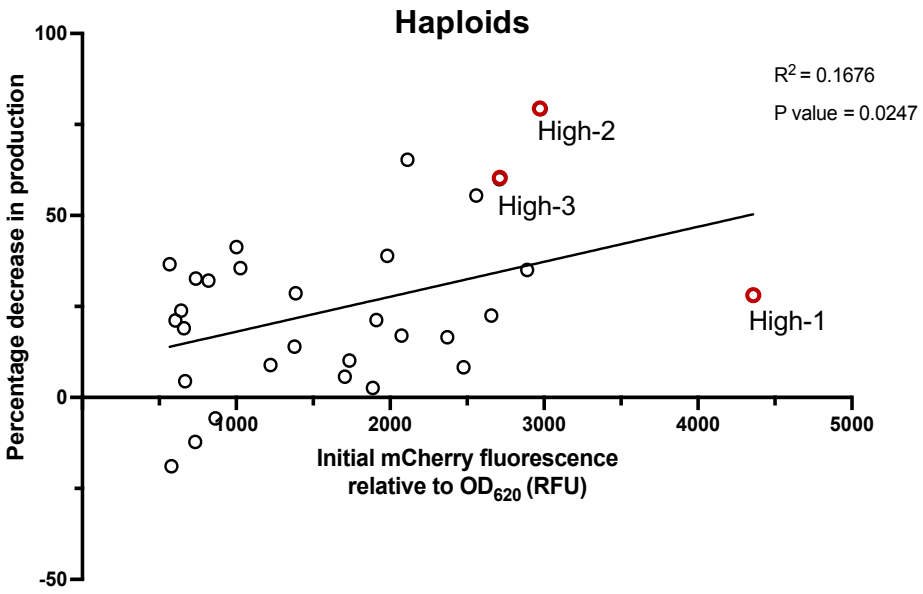
Figure 3.5: SARS-CoV-2 spike fragment mCherry fusion protein expression over time in selected strains from the haploid and diploid libraries. mCherry level was measured and normalized to OD_{620} every ~10 generations during sequential subculture into fresh medium every 3 days. Shown are expression profiles over 140 generations for three high-producing strains (High-1, High-2, High-3) and three low-producing strains (Low-1, Low-2, Low-3) from the haploid (A) or diploid (B) library expressing SARS-CoV-2-mCherry fusion protein. Each data point represents the mean value ± SEM of eight biological replicates (i.e. eight clones derived from a single strain stock).

3.3.2 Correlation between initial production level and production decline

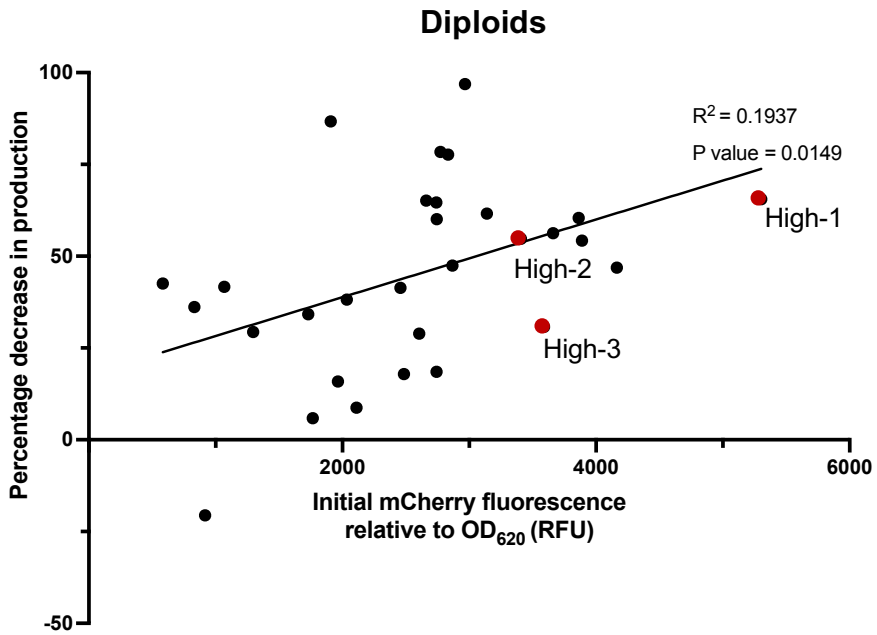
To test whether there is a correlation between the initial production level and production stability amongst genetically diverse strains, 24 additional strains were selected from each of the haploid and diploid libraries expressing the SARS-CoV-2-mCherry fusion protein to represent a broad range of expression levels. These strains were cultivated in 48-well microtiter plates and monitored for mCherry expression over 140 generations, under the same serial passaging conditions used previously. To quantify production stability, the percentage decrease in mCherry expression over 140 generations was calculated using data from both the 24 additional strains and the six previously analysed high- and low-producers (**Figure 3.5**). The results revealed a moderate negative correlation between initial production level and the extent of production decline over 140 generations (**Figure 3.6**): the strains with higher initial production levels generally experienced the greater declines in production over time. This trend was observed in both haploid and diploid backgrounds. While the results

suggested that higher production imposes more burden and contributes to instability, the relationship was not absolute. For the combined dataset of 60 strains, the coefficient of determination (R^2) was 0.2747, indicating that approximately 27.5% of the variation in long-term production decline could be explained by the initial expression level.

A


Haploids
R2 = 0.1676
P value = 0.0247
High-2
High-3
High-1
Percentage decrease in production
100
50
0
-50
1000
2000
3000
4000
5000
Initial mCherry fluorescence
relative to OD620 (RFU)

B


Diploids
R2 = 0.1937
P value = 0.0149
High-1
High-2
High-3
Percentage decrease in production
100
50
0
-50
2000
4000
6000
Initial mCherry fluorescence
relative to OD620 (RFU)

C

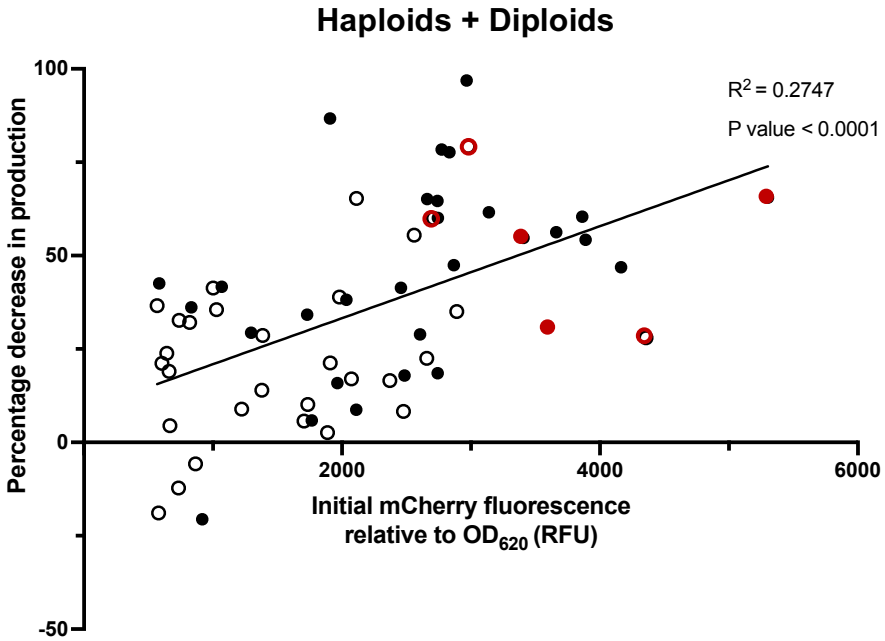

Haploids + Diploids
R2 = 0.2747
P value < 0.0001
Percentage decrease in production
100
50
0
-50
2000
4000
6000
Initial mCherry fluorescence
relative to OD620 (RFU)

Figure 3.6: Correlation between initial SARS-CoV-2-mCherry fusion protein production level and the magnitude of production decline after 140 generations.

Strains were cultured in 48-well plates for 72 h in BMMD + CSM – Leu – Met medium. (A, B) Data represent 30 strains (24 randomly selected strains plus the six previously analysed in **Figure 3.5**) from the haploid (A) and diploid (B) libraries, respectively. (C) Combined data for all 60 strains. Each data point represents the mean value from eight biological replicates (i.e. clones derived from a single strain), with haploid strains shown as open circles and diploid strains shown as filled circles. The three selected high-producing strains from each library (shown in **Figure 3.5**) are highlighted in red.

3.3.3 Natural diversity enables the identification of robust and durable high-producing strains

While a general trend was observed linking higher initial production levels with increased production decline, some strains clearly constituted outliers to this rule (**Figure 3.6**). For example, the High-1 strain from the haploid library, which showed by far the highest initial production level among the 30 haploid strains tested, maintained moderate to high production throughout the 140 generations (**Figure 3.5 and Figure 3.6A**). In contrast, the High-2 and High-3 haploid strains, which had relatively lower initial production levels than High-1, exhibited substantial declines in production and eventually reached levels comparable to those of the initial low producers (**Figure 3.5**). A similar pattern was observed among the selected diploid strains. The High-2 and High-3 diploid strains exhibited comparable initial production levels, but their production stability differed markedly: the High-2 diploid strain experienced an approximately 50% decrease in production over 140 generations, whereas the High-

3 diploid strain showed only about a 25% decrease (**Figure 3.6B**). These results highlight the functional diversity present within the starting strain pool, where some rare genotypes may possess both high expression capacity and resilience to production-associated stresses.

3.4 Discussion

3.4.1 Secreted proteins impose a greater production burden

This chapter compared the long-term production dynamics of two heterologous proteins, an intracellular AP205 VLP-mCherry and a secreted SARS-CoV-2 spike fragment-mCherry fusion, across genetically diverse *S. cerevisiae* strains from the libraries. The high-producing strains expressing the SARS-CoV-2-mCherry protein exhibited significant declines in production over 140 generations, whereas those expressing the AP205 VLP-mCherry protein remained relatively stable. These findings support the hypothesis that secreted proteins impose a greater metabolic burden than intracellular ones, likely due to the additional demands on the secretory pathway, trafficking machinery and intracellular redox homeostasis [175; 176]. The relatively stable expression of AP205 VLP-mCherry, observed in both high- and low-producing strains, implies that cytosolic accumulation of this protein did not impose sufficient fitness cost to affect production over time. In contrast, secretion-associated stresses in the high-producing strains for SARS-CoV-2-mCherry appear to have driven evolutionary selection for low-producing variants with better growth fitness and triggered greater instability. These findings highlight how protein-specific properties, such as whether the protein is secreted or retained intracellularly, can influence cellular burden and evolutionary trajectories. However, it should be noted that each strain was derived from an individual colony from the libraries; so differences in genetic background might also contribute to the observed effects. A more controlled comparison between intracellular and secreted expression of the same protein in an isogenic background could strengthen this conclusion. Nonetheless, such a direct comparison was beyond the scope of the current study. The primary aim here was to identify a suitable model protein for studying evolutionary dynamics in yeast

biomanufacturing systems. Based on its pronounced production decline, the SARS-CoV-2-mCherry fusion protein was chosen as a more appropriate model for investigating production instability and population shifts in yeast-based bioproduction.

3.4.2 Multiple determinants of long-term production stability

Although a general trend was observed linking higher expression with greater production decline, the modest correlation indicates that initial production alone does not fully explain the production decline. Other factors must contribute to the remaining variation observed across strains. One important factor is strain-specific variation in stress tolerance. Recombinant protein expression, especially of secreted proteins, can activate cellular stress responses, such as the unfolded protein response and oxidative stress pathways, which impair protein homeostasis and reduce cellular fitness over time [158]. Yeast strains differ in their capacity to manage these stresses. For example, a comparative study across twelve *S. cerevisiae* strains revealed marked differences in stress responses under glucose limitation, oxidative stress, osmotic stress and acid stress [177], suggesting significant variation in stress tolerance across strains. Yeast strains with weaker stress tolerance may, therefore, be more prone to productivity loss under sustained burden, regardless of their initial expression levels.

In particular, differences in plasmid stability are also a likely contributor that may underlie long-term production differences. Genetic background has been shown to influence 2-micron plasmid retention in *S. cerevisiae*. A well-characterised example is the gene *MMS21*, which encodes a SUMO E3 ligase that is a component of the Smc5/6 complex, involved in DNA repair, chromosome segregation and sister chromatid cohesion [178-180]. Natural variation at the *MMS21* locus has been linked

to differences in 2-micron plasmid stability, with one variant promoting more efficient plasmid retention than another [180]. As a result, strains with genetic backgrounds that favour plasmid maintenance may exhibit more stable protein expression over time, even when their initial production levels are comparable to those of strains with less stable plasmid inheritance.

3.4.3 Harnessing natural diversity for selection and engineering of robust strains

The multiple determinants of long-term production stability discussed above underscore the importance of host genomic context in shaping expression dynamics. The observed variation among the 60 selected strains suggests that it is feasible to identify genotypes that combine high productivity with stable expression. The presence of such ‘stay-high’ strains, such as the High-1 haploid strain and the High-3 diploid strain identified in this chapter, demonstrates that it is indeed possible to achieve more durable productivity under prolonged biomanufacturing conditions. Their existence emphasizes the value of screening large-scale, genetically diverse strain libraries to identify not only high producers, but also those that remain stable over time. These strains likely harbour favourable combinations of traits such as stress tolerance, efficient plasmid maintenance and metabolic balance that mitigate the negative consequences of sustained heterologous protein production. Beyond direct use, these naturally robust strains can serve as starting points for further improvement. Strategies such as mutagenesis, breeding, and directed evolution can be used to generate even greater diversity and refine desirable traits [45], ultimately enabling the development of superior industrial yeast strains tailored for specific production tasks. Additionally,

these strains can be used as a high-performance reference pool for genetic dissection. For example, quantitative trait locus (QTL) mapping among ‘stay-high’ strains could help identify genomic regions associated with production stability [69], offering valuable insights into the genetic basis of robustness in industrial yeast.

3.4.4 Conclusions

This chapter demonstrated that production burden and natural strain diversity can both shape the stability of heterologous protein expression in *S. cerevisiae* during extended cultivation. By comparing two model proteins, an intracellular AP205 VLP-mCherry and a secreted SARS-CoV-2-mCherry, it was evident that protein-specific features also affect production stability. A moderate correlation was observed between initial production level and production decline for SARS-CoV-2-mCherry-producing strains, suggesting that high-yield production can increase selection pressure favouring low-producing variants. Importantly, differences among strains with similar starting yields revealed that certain genotypes can combine high productivity with greater stability. This highlights the value of exploiting natural diversity to identify robust production hosts. Together, the findings underscore that production decline is shaped by multiple, interacting factors, including expression burden, host stress tolerance and plasmid stability, all of which are influenced by the genetic background of the host. Exploiting this natural diversity through large-scale screening and targeted improvement strategies offers a practical route to developing more stable, efficient yeast strains for industrial biomanufacturing.

4 GENETIC DRIVERS OF PRODUCTION INSTABILITY IN HIGH-PRODUCING YEAST

4.1 Introduction

As described in Chapter 1, the global biopharmaceutical market size is over 500 billion USD, with recombinant proteins accounting for more than 200 billion USD of this total [10]. This market continues to expand rapidly at an annual growth rate of 8.5%, and yeast remains an important production host, responsible for around one-sixth of all biopharmaceutical proteins for human use [10; 52; 57; 181]. Of these, *S. cerevisiae* produces the vast majority of FDA-approved biologics from yeast. The production decline observed in high-producing yeast strains expressing the SARS-CoV-2-mCherry fusion protein over 140 generations, as described in Chapter 3, highlights a key biological challenge that can impact the cost-effectiveness and reliability of industrial biomanufacturing processes. Therefore, understanding the underlying causes of this decline could enable the development of more robust yeast strains capable of maintaining stable, high-yield production over extended cultivation. This would lower production costs by reducing the frequency of seed train cultures and allowing continuous culture or extended fill and draw fermentations campaigns.

Cell population heterogeneity, introduced in Chapter 1, is a factor that can drive production instability in microbial bioprocesses. This heterogeneity between individual cells can arise from both genotypic and phenotypic sources [22]. Genotypic heterogeneity can arise from events such as plasmid loss and mutations in either the production plasmid or the host genome that reduce or abolish expression in cell sub-populations, any of which could dominate the culture over time [22; 82; 85]. For

example, plasmid loss may arise through random integration of the selectable marker into the genome, thereby relieving the requirement to maintain the plasmid under selection pressure [182]. Mutations within the production plasmid can also compromise production stability. For instance, point mutations or deletions in the promoter region can reduce or abolish transcriptional activity [183], while insertions or deletions in the coding sequence can introduce frameshifts that produce truncated or misfolded proteins, usually targeted for degradation by cellular quality control mechanisms such as the proteasome and lysosome in yeast [184-186]. Similarly, mutations in chromosomal genes involved in transcriptional regulation or nutrient uptake can also contribute to production instability. For instance, mutations in transcriptional regulators that control the promoter activity driving heterologous expression [187]. In contrast, phenotypic heterogeneity involves non-genotypic, reversible differences among genetically uniform cells, such as in plasmid copy number or arising from gene expression noise [96; 188]. Within a genetically uniform cell population, low-producing cells with lower plasmid number or reduced expression burden may suffer less stress (e.g. from metabolic demand) and therefore gain a growth advantage over high-producing cells. Both genotypic and phenotypic heterogeneity can be subject to selection during extended cultivation, favouring low- or non-producing variants with reduced metabolic burden. Over time, this selection leads to enrichment of those low and/or non-producing variants in the population and ultimately results in an overall decline in production yield [22; 83; 145].

In this thesis, recombinant protein production relied on a whole-2-micron plasmid-derived expression system. Episomal plasmids based on the native yeast 2-micron plasmid are commonly used in biotechnology because of their high copy number,

resulting in elevated product yields [189]. These plasmids are typically maintained at 40-80 copies per haploid genome, making it attractive for overexpression of a product gene [68; 190-192]. In whole-2-micron plasmids, this high-copy maintenance depends on specific genetic features on the 2-micron-derived plasmid: the FLP recombinase gene and FLP recognition target (*FRT*) sites mediate site-specific recombination and amplification, while the *REP1*, *REP2* and *RAF1* genes form a partitioning system that ensures equal segregation of plasmid copies to daughter cells [67; 193; 194]. Together, these mechanisms maintain copy number and inheritance fidelity, supporting sustained high-level expression under selective conditions. However, the large number of plasmid copies also expands the mutational target size, increasing the chance of adaptive mutations that might reduce production burden over time [195]. For example, mutations in genes responsible for amplification or partitioning may lead to decreased copy number and reduced production levels.

The stability of whole-2-micron-derived plasmids is also influenced by chromosomal factors. Host-encoded genes can affect plasmid replication, partitioning and amplification. A typical example is *MMS21*, an essential gene encoding a SUMO E3 ligase component of the Smc5/6 complex, which plays important roles in DNA repair, chromosome segregation and sister chromatid cohesion [178-180]. Natural sequence variation in *MMS21* has been shown to impact 2-micron plasmid stability: a single amino acid substitution (Thr69Ile) was identified as a causal variant in a QTL mapping study. Strains carrying the Ile69 allele exhibited relatively lower 2-micron plasmid stability compared to those with the Thr69 allele [180]. Importantly, while plasmids (especially high-copy-number plasmids) can be lost, segregate unevenly, or be replaced by other plasmid variants within a single cell, chromosomal DNA is faithfully

replicated once per division, with more reliable mechanisms ensuring accurate partitioning and inheritance by all descendants [196].

4.1.1 Aims

It remained unclear whether the production decline observed in the selected high-producing SARS-CoV-2-mCherry strains of Chapter 3 was driven by accumulated genetic changes and/or shifts in phenotypic heterogeneity. This chapter aims to investigate the underlying causes of this production decline. To address this, evolved cultures of the selected high-producing SARS-CoV-2-mCherry strains at the 140th generation were analysed.

Retransformation assays were performed to determine whether the decline could be due to changes in the production plasmid or chromosome. Whole plasmid sequencing was carried out to identify mutations that may have accumulated in the production plasmid during extended cultivation. The functional consequences of these mutations on production stability were also assessed. The contribution of natural variation in the *MMS21* gene to production stability, specifically the Thr69 and Ile69 alleles, was also examined.

4.2 Chapter-specific materials and methods

4.2.1 Yeast strains and culture conditions

In this chapter, the 30 yeast strains selected from the haploid library described in Chapter 3 were used, along with the selected haploid strains High-1, High-2, High-3 at the beginning (10th generation) and the end (140th generation) of the long-term production experiment (Chapter 3). Unless otherwise specified, these strains and their derivatives were maintained and grown in BMMD + CSM – Leu – Met medium and were generally cultured in 48-well microtiter plates with 500 µL medium per well at 30 °C, 200 rev. min^{-1} (50 mm orbit). For repression of production, BMMD + CSM – Leu – Met medium was supplemented with methionine to a final concentration of 3 g/L (20 mM) to make BMMD + CSM – Leu + Met medium.

For assays of the *MET17* promoter activity and the FLP/*FRT*-mediated recombination efficiency (below), the haploid laboratory *S. cerevisiae* strain BY4741 (*MAT***a**; *his3*Δ*1*; *leu2*Δ*0*; *met15*Δ*0*; *ura3*Δ*0*) was used. Wild-type BY4741 and its engineered derivatives were maintained and grown in YNB medium supplemented as required with amino acids and/or nucleotide bases, or in YPD medium. The cultures were generally grown in 10 mL medium within 50 mL Erlenmeyer flasks, inoculated from single colonies and incubated at 30 °C, 120 rev. min^{-1} (50 mm orbit).

4.2.2 Curing of the production plasmid from yeast

The selected haploid strains were cured of the production plasmid by serial passaging in non-selective medium. Briefly, each strain was cultured in 500 µL YPD medium in 48-well microtiter plates at 30 °C, 200 rev. min^{-1} (50 mm orbit) overnight. An aliquot

(10 µL) of the overnight culture was transferred into fresh YPD medium and incubated under the same conditions. This passaging step was repeated for five cycles to promote plasmid loss. After the final passage, the overnight cultures (OD_{600} ~1) were harvested by centrifugation (1000 x *g,* 2 min) and diluted to OD_{600} ~ 10^{-4}. Aliquots (100 µL) of diluted samples were spread onto YPD agar plates and incubated at 30 °C for two days. Arising colonies (10-15 per strain) were picked and patched onto both BMMD + CSM – Leu – Met agar plates and YPD agar plates. The plates were incubated at 30 °C for two days. Colonies growing only on YPD plates but failing to grow on BMMD + CSM – Leu – Met plates were identified as plasmid-cured strains.

4.2.3 Plasmid extraction from yeast

The production plasmids were extracted directly from yeast cells, following the method described by Singh and Weil [197], with minor modifications made specifically for this project in the volumes of cultures and reagents used. Briefly, yeast cultures (100 mL) were grown in BMMD + CSM – Leu – Met medium in 500 mL Erlenmeyer flasks from single colonies at 30 °C, 200 rev. min^{-1} (50 mm orbit) overnight. The Qiagen Plasmid Miniprep Kit (Qiagen) was used for plasmid extraction, according to the manufacturer's instructions, with the following modifications: cells were harvested by centrifugation (1000 x *g*, 5 min) and resuspended in 2 mL Buffer P1 by vortexing. The cell suspension was mixed with 80 µL of 75 units/ µL lyticase from *Arthrobacter luteus* (Sigma) and incubated at 37 °C for 2 h to digest the cell wall. Cells were then lysed by adding 2 mL Buffer P2 and mixed by gentle inversion, followed by a 10-minute incubation at room temperature. The lysate was neutralised by adding 2.8 mL of Buffer N3 and mixed gently by inverting, then incubated for 30 min on ice. The supernatant was collected by centrifugation (13,000 x *g*, 10 min) at 4 °C and loaded onto three Qiagen plasmid

miniprep columns to maximise recovery from the larger culture volume. The columns were washed once with 500 µL Buffer PB and then washed twice with 750 µL Buffer PE. Plasmid DNA was eluted by adding 30 µL nuclease-free water to each column, incubating at room temperature for 5 min and centrifuging at 13,000 x *g* for 1 min.

4.2.4 Whole plasmid sequencing

Plasmid DNA concentrations were determined using the Qubit 4 Fluorometer (Invitrogen), with the Qubit dsDNA High Sensitivity Assay Kit (Invitrogen), according to the manufacturer's protocols. Plasmid samples were prepared at a concentration of 50-100 ng/µL in 10 µL of nuclease-free water and then submitted to Full Circle Labs for whole plasmid sequencing using Oxford Nanopore technology.

4.2.5 Sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

For SDS-PAGE analysis, retransformants containing the single-type plasmids (10 mL cultures) were grown in BMMD + 2xCSM – Leu – Met medium (with double the standard amount of CSM, i.e. 0.138% w/v) in 50 mL Erlenmeyer flasks at 30 °C, 120 rev. min^{-1} (50 mm orbit) for three days. After three days, 1 mL of culture was centrifuged at 10,000 x *g* for 2 min, and 500 µL of the resulting supernatant was collected. A 32.5 µL aliquot of each supernatant sample was mixed in a PCR tube with 5 µL of NuPAGE™ Sample Reducing Agent (10x, ThermoFisher) and 12.5 µL of NuPAGE™ LDS Sample Loading Buffer (4x, ThermoFisher). The mixtures were heated at 80 °C for 10 min on a heat block to denature the proteins. Prepared samples (30 µL) were then loaded onto pre-cast NuPAGE™ Bis-Tris gel (4-12%, ThermoFisher) in an Xcell SureLock™ electrophoresis chamber filled with 1x MES SDS Running

buffer, prepared from NuPAGE™ MES SDS Running Buffer (20x, ThermoFisher), with 500 µL of NuPAGE™ Antioxidant (ThermoFisher) added to the inner chamber. 5 µL of PageRuler™ Prestained Protein Ladder (10 to 180 kDa, ThermoFisher) was used as a molecular weight marker. SDS-PAGE was run at 200V for 35 min. Once the dye front had fully migrated, the gel was removed and stained using GelCode™ Blue Safe Protein Stain (ThermoFisher) according to manufacturer's instructions.

4.2.6 *MET17* promoter activity assay

Reporter plasmids were constructed to assess *MET17* promoter activity. Primers used were listed in **Table Appendix B.2**. The wild-type *MET17* promoter (wt*pMET17*) and mCherry coding sequence were amplified from the production plasmid extracted at the 10th generation. The mutated *MET17* promoter (mu*pMET17*) was amplified from the production plasmid carrying the identified mutation, and the *CYC1* terminator (*tCYC1*) was amplified from the pCM190 plasmid (Euroscarf). These fragments were assembled between the *Sac*I and *Sal*I restriction sites in pRS315 to generate the reporter constructs pRS315-wt*pMET17-mCherry-tCYC1* and pRS315-mu*pMET17-mCherry-tCYC1*, via Gibson assembly (see **Section 2.4.3**). The assembled reporter plasmids were used to transform *S. cerevisiae* strain BY4741. Transformations were performed using the standard procedure as described in General Materials and Methods 2.3.5. Transformants were selected on YNB agar plates supplemented with 0.02 mg/mL uracil, 0.02 mg/mL methionine and 0.02 mg/mL histidine to select for leucine prototrophy conferred by the *LEU2* marker of pRS315. Four transformants for each transformation were picked and grown overnight in YNB broth supplemented with 0.02 mg/mL uracil, 0.02 mg/mL methionine and 0.02 mg/mL histidine. Aliquots

(500 µL) of the overnight cultures were transferred to 48-well plates for mCherry measurement as described in **Section 2.5**.

4.2.7 Quantification of plasmid copy number by droplet digital PCR (ddPCR)

Yeast cultures (10 mL) were grown in BMMD + CSM – Leu – Met medium from single colonies in 50 mL Erlenmeyer flasks at 30 °C, 120 rev. min^{-1} (50 mm orbit) overnight. An aliquot (1.5 mL) of the overnight culture was harvested by centrifugation (1000 x *g*, 5 min) and used in the following extraction procedure. Total DNA was extracted from yeast cells using the DNeasy Blood & Tissue Kit (Qiagen), following the manufacturer's supplementary protocol for purification of total DNA from yeast. DNA concentration was measured using the Qubit 4 Fluorometer, with the Qubit dsDNA High Sensitivity Assay Kit, and then diluted to a concentration of 0.1 ng/ µL in nuclease-free water. For ddPCR, 2 µL of diluted DNA (0.1 ng/µL) was used as a template. Reactions were prepared in a 96-well QIAcuity Nanoplate 8.5k (Qiagen) with a total reaction volume of 12 µL, using the QIAcuity EvaGreen (EG) PCR Kit (Qiagen), according to the manufacturer's instructions. Each 12 µL reaction included 1 µL of *Sac*I-HF restriction enzyme (NEB) to linearize the template DNA, with digestion performed by incubating the plate at room temperature for 10 min. After the 10-min incubation, the reaction was run on the QIAcuity One Digital PCR System (Qiagen), with the following thermal cycling conditions: 95 °C for 2 min for 2 min initial activation, followed by 40 cycles of 95 °C for 15 sec (denaturation), 55 °C for 15 sec (annealing) and 72 °C for 15 sec (extension), with a final cooling step at 40 °C for 5 min and an extended hold at 10 °C. Data were analysed using QIAcuity Software Suite 1.2.18. The single-copy *S. cerevisiae ACT1* housekeeping gene was used as a reference,

and a plasmid-specific noncoding sequence was used as the target for plasmid copy number quantification. Plasmid copy number per genome was determined by normalizing plasmid target concentration (copies/μL) to *ACT1* concentration (copies/μL).

4.2.8 Construction of engineered *S. cerevisiae* BY4741 strain expressing FLP recombinase

To assess the efficiency of FLP/*FRT*-mediated recombination with different *FRT* site variants, *S. cerevisiae* BY4741 was engineered to express FLP recombinase. This was achieved by integrating the *FLP* gene under the control of the inducible *GAL1* promoter (*pGAL1*) at the *HO* locus, using HIS3MX6 as a selection marker. Primers used were listed in **Table Appendix B.2**. To do this, the *FLP* open reading frame (ORF) was amplified from the production plasmid extracted from yeast at 10th generation. A *tCYC1* sequence was amplified from plasmid pCM190. The *FLP* ORF and *tCYC1* were assembled between the *Sac*I and *Hind*III restriction sites in pFA6a-*HIS3MX6-pGAL1* to give pFA6a-*HIS3MX6-pGAL1-FLP-tCYC1* construct, via Gibson assembly as described in **Section 2.4.3**. The *HIS3MX6-pGAL1-FLP-tCYC1* fragment was amplified by PCR with 50 bp homology arms targeting the *HO* locus and then used for transformation into *S. cerevisiae* BY4741 as described in 2.4.5. Transformants (*HO*Δ::*His3MX6-pGAL1-FLP-tCYC1*) were selected in YNB agar supplemented with 0.02 mg/mL uracil, 0.1 mg/mL leucine and 0.02 mg/mL methionine. Correct genomic integration was confirmed by PCR and gel electrophoresis using genomic DNA extracted from the transformants as template.

4.2.9 FLP/*FRT*-mediated recombination efficiency assay

Primers used were listed in **Table Appendix B.2**. *FRT* site variants to be tested were cloned in the same orientation flanking a *KanMX* resistance cassette, using plasmid pRS315 as a backbone, such that recombination by the FLP recombinase would excise the cassette rather than invert it. The *FRT-KanMX-FRT* fragments were amplified by PCR using overhang primers containing *Sac*I and *Sal*I restriction sites together with desired *FRT* sequences. The PCR product was digested with *Sac*I and *Sal*I restriction enzymes (NEB) and ligated into pRS315 using T4 DNA ligase (NEB), according to the manufacturer's protocols. The resulting plasmid constructs were transformed into the engineered BY4741 strain expressing FLP recombinase. Transformants were selected on YNB agar plates supplemented with 0.02 mg/mL uracil and 0.02 mg/mL methionine.

For the recombination assay, transformants carrying *FRTs*-flanked *KanMX* constructs were first precultured in 10 mL YNB broth supplemented with 0.02 mg/mL uracil and 0.02 mg/mL methionine to maintain selection, with 2% glucose as the carbon source, in 50 mL Erlenmeyer flasks at 30 °C, 120 rev. min^{-1} (50 mm orbit) overnight. Cells (OD_{600} ~ 1) were then harvested by centrifugation (1000 x *g,* 2 min) and resuspended to OD_{600} ~ 0.1 in 10 mL YNB broth supplemented with 0.02 mg/mL uracil and 0.02 mg/mL methionine but with 2% galactose, instead of glucose, as the carbon source to induce *FLP* expression. Induced cultures were incubated at 30 °C, 120 rev. min^{-1} (50 mm orbit), with serial passaging (diluting overnight cultures at OD_{600} ~ 3-4 to OD_{600} ~ 0.1 in fresh selective YNB medium with 2% galactose as carbon source) for approximately 50 generations to maintain continuous *FLP* induction. After 50 generations, cultures were diluted to OD_{600} ~ 1 x 10^{-4} and 100 µL aliquots were plated

onto non-selective YPD agar and YPD agar containing 200 μg/mL G418. Plates were incubated at 30 °C for 2 days, and colonies were counted. FLP/*FRT*-mediated recombination efficiency was calculated following the equation:

$$\frac{(\text{the number of colonies on YPD} - \text{the number of colonies on YPD containing G418})}{\text{the number of colonies on YPD}}$$

4.3 Results

4.3.1 Differential recovery patterns emerged among evolved strains following repression of production

To investigate the causes of SARS-CoV-2-mCherry production declines observed in Chapter 3, haploid strains were chosen for further analysis because their single set of chromosomes simplifies genetic interpretation and analysis compared to diploid strains, where, for example, recessive mutations are commonly masked. Additionally, haploids are generally easier to manipulate [198]. The haploid High-1, High-2 and High-3 strains from the long-term production experiment described in Chapter 3 (**Figure 3.5A**) were examined.

Phenotypic heterogeneity in microbial populations can arise through phenotypic switching, a process that generates subpopulations better adapted to specific environmental conditions [199]. In the context of recombinant protein production, cells experiencing a high production burden may tend to downregulate production to improve fitness. If such a production decline is primarily driven by a process like phenotypic switching, temporarily repressing expression might enable cells to switch back to a production-competent state when stress is relieved. To test this hypothesis, the eight evolved cultures of each starting strain (maintained under production-inducing conditions for 140 generations) were grown in BMMD + CSM – Leu + Met medium for approximately 20 generations to repress expression from the *MET17* promoter. This repression was intended to reduce production-associated burden and allow cells to recover. After this repression phase, cultures were transferred back into BMMD + CSM – Leu – Met medium to re-induce production and SARS-CoV-2-mCherry production was measured (**Figure 4.1**).

For the High-1 strain, no significant increase in production was observed after the repression and re-induction, with expression levels remaining comparable to those observed at the 140th generation. This suggests that for High-1 cultures, the production declines were largely fixed and unresponsive to short-term changes in induction conditions. In contrast, both High-2 and High-3 strains displayed significant increases in average production across their eight evolved cultures after the repression phase compared to their production levels at the 140th generation. Nevertheless, the production levels after repression and re-induction in these strains remained significantly lower than the initial levels. Notably, among the High-2 cultures, there were marked differences in response between cultures. Specifically, evolved cultures 1, 2, 4 and 7 showed evident improvement in production after repression and re-induction, while cultures 3, 5, 6 and 8 remained at low production levels (**Figure 4.1**). This suggested that over 140 generations, replicate cultures of the same starting strain diverged through different evolutionary trajectories. Some cultures likely acquired genetic changes that fixed the low production phenotype, making them unresponsive to the repression, while others may have retained or developed mixed strategies, potentially combining phenotypic adaptation with genetic changes.

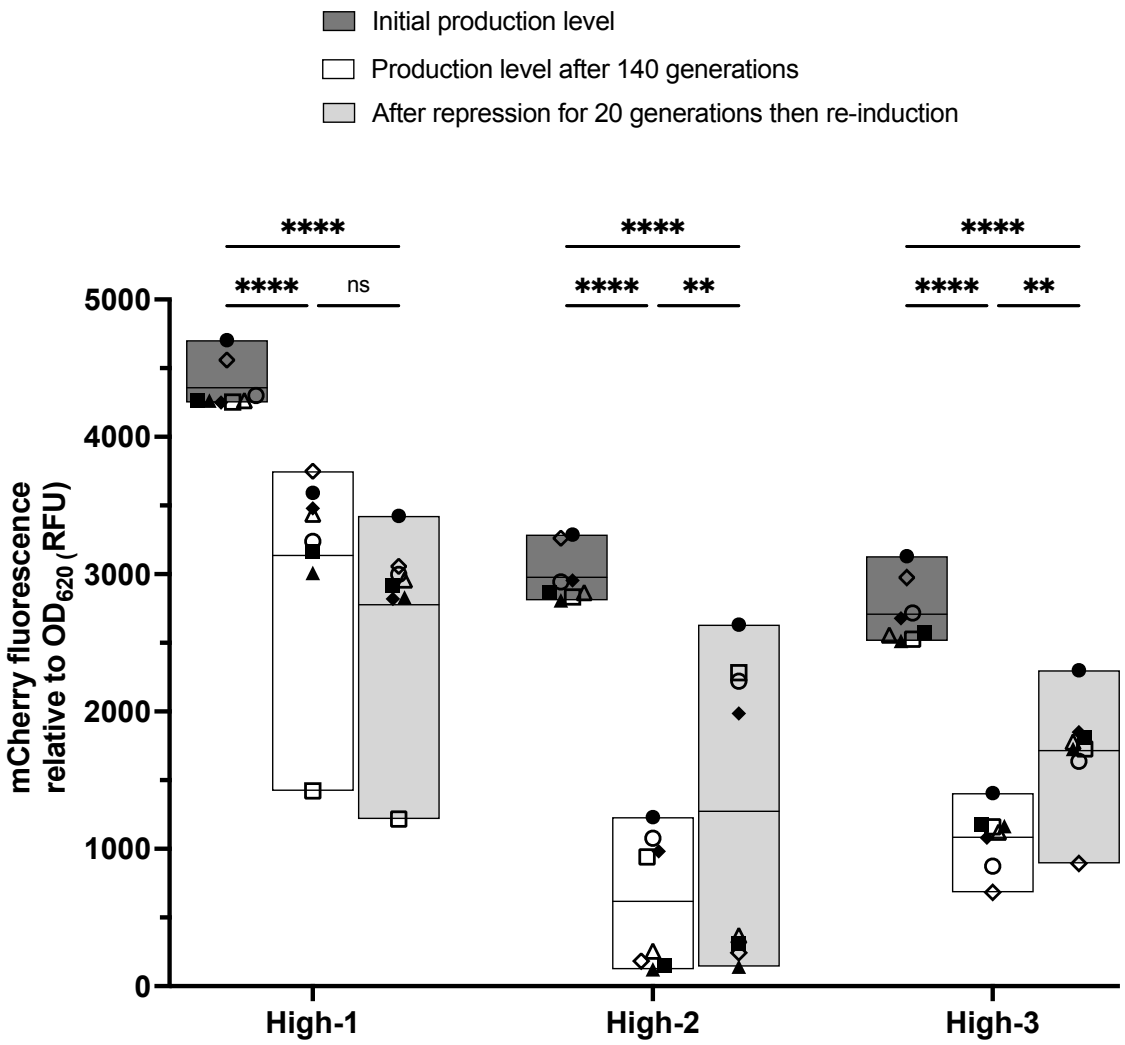


Figure 4.1: Expression of the SARS-CoV-2-mCherry fusion protein from evolved cultures following repression of production and re-induction. Initial production levels and production levels after 140 generations are shown as controls. For each strain, eight evolved cultures previously passaged for 140 generations under production conditions were grown in 48-well microtiter plates with 500 µL BMMD + CSM – Leu + Met medium for approximately 20 generations (two sequential 3-day passages of ~10 generations each) to repress *MET17* promoter activity. After this repression phase, cultures were transferred into 500 µL BMMD + CSM – Leu – Met medium to re-induce production. mCherry levels were measured and normalised to OD_{620}. The control data are shown for the three haploid strains High-1, High-2 and

High-3 from **Figure 3.5A**. Symbols indicate the eight replicate cultures originally established for each strain in the long-term production experiment: culture 1 (●), culture 2 (○), culture 3 (■), culture 4 (□), culture 5 (▲), culture 6 (△), culture 7 (◆), culture 8 (◇). The floating bars represent the full data range (min to max), with a line indicating the mean. Statistical analysis (Two-way ANOVA): $F(4, 42) = 6.627$, $p = 0.0003$. Significance is indicated as $p < 0.01$ (**), $p < 0.0001$ (****), not significant (ns).

4.3.2 Selection of candidate colonies for further analysis

To investigate potential genetic factors underlying the fixed low-production phenotype observed in evolved cultures, the haploid High-2 and High-3 strains were analysed in greater detail, as these showed the most significant production declines in Chapter 3 (**Figure 3.5A**). For each of these two strains, three evolved cultures with the lowest production at the 140th generation were selected for further study (High-2_culture 3, 5, and 8 and High-3_culture 2, 7 and 8). These evolved cultures originated from eight replicate cultures initially established for each strain in the long-term production experiment, but they could no longer be considered as clones after the long-term cultivation, as they likely contained genetically diverse cells following mutation and selection within each culture. To select candidate colonies for further analysis, cells from each evolved culture were plated out on BMMD + CSM – Leu – Met agar plates. Sixteen arising colonies derived from each evolved culture were randomly picked and cultured separately in BMMD + CSM – Leu – Met liquid medium for three days. mCherry level was then measured to assess recombinant protein production levels at single-colony resolution (**Figure 4.2**). Colonies derived from the High-2 cultures generally exhibited uniformly low expression, with only one colony from culture 3

showing markedly higher production. In contrast, colonies derived from High-3 cultures displayed greater variation in production within the set of 16 selected colonies from each culture. Additionally, the production levels of all colonies from High-3 cultures were generally higher than those from High-2 cultures (excluding the single High-2_culture 3 colony that had elevated expression). These results suggest that High-2 cultures might have evolved more rapidly toward a stable, low-production phenotype. The greater variability among colonies from High-3 cultures indicates that those cultures may still have been in the process of evolving toward lower production levels. It also suggests that High-3 might be relatively more stable under production conditions than High-2, potentially maintaining higher expression for longer before fully adopting a low-production state. From each evolved culture, the colony with the lowest expression (indicated by red datapoints in **Figure 4.2**) was selected as the candidate for further detailed genetic analyses.

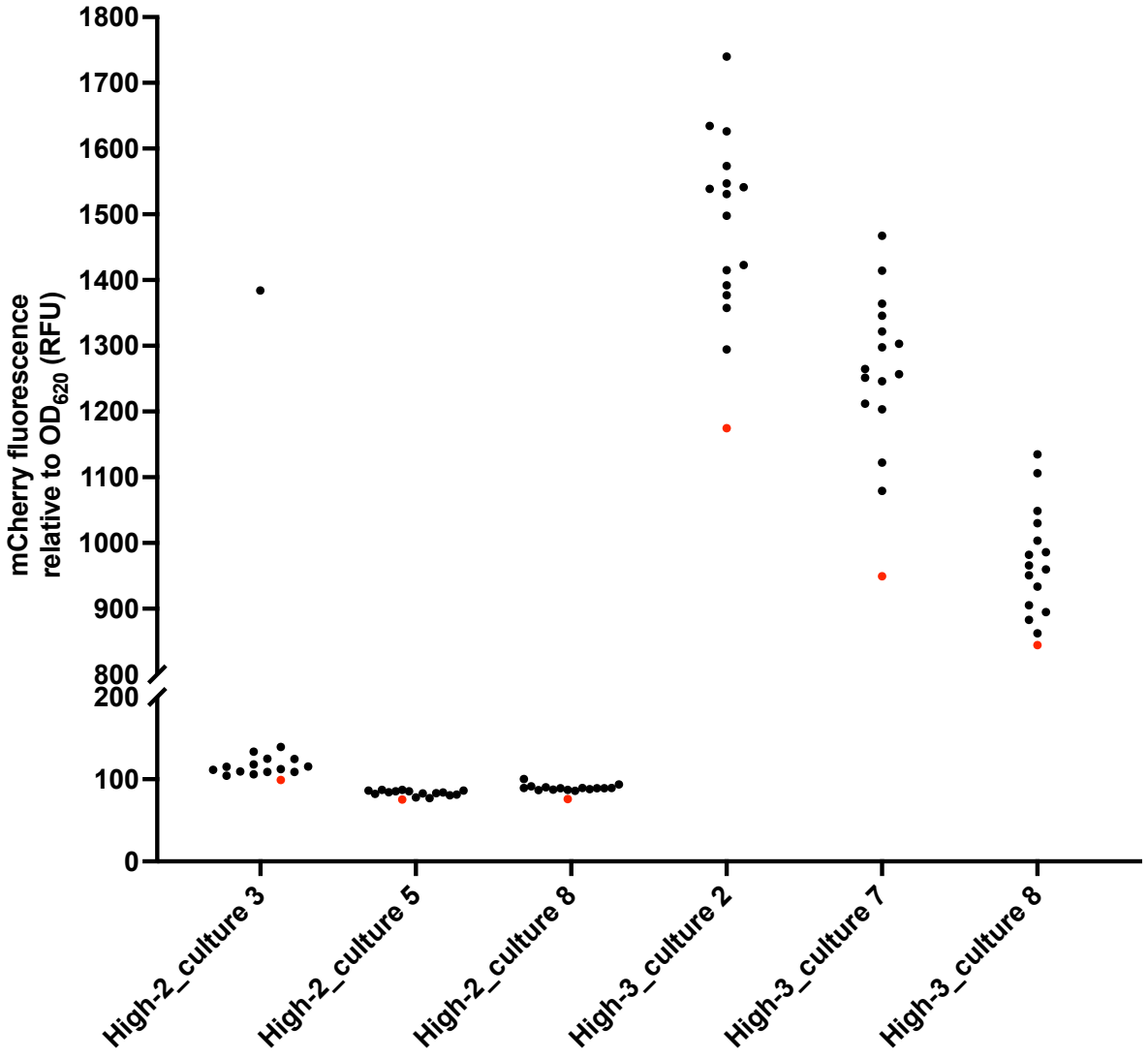


Figure 4.2: Expression of the SARS-CoV-2-mCherry fusion protein from individual colonies isolated from low-producing evolved cultures after production for 140 generations. Sixteen individual colonies from each evolved culture were grown in 48-well microtiter plates in 500 µL BMMD + CSM – Leu – Met medium for 3 days, before measuring cellular mCherry fluorescence with normalisation to OD_{620}. Shown are the SARS-CoV-2-mCherry fusion protein production levels of each of the 16 colonies isolated from each of the three lowest-expressing High-2 evolved cultures (culture 3, 5 and 8) and the three lowest-expressing High-3 evolved cultures (culture 2, 7 and 8) after 140 generations. Each

data point is the mean of approximately 1.5 x 10^8 cells analysed. Red data points indicate colonies selected for subsequent retransformation and plasmid analysis tests.

4.3.3 Strain curing and retransformation showed that production decline was primarily due to plasmid-borne factors

To determine what proportion of the production decline observed in long-term production experiment was due to changes in the genome and/or in the production plasmid, selected low-expressing cultured colonies (red datapoints in **Figure 4.2**) were first cured of their production plasmids and retransformed with production plasmids extracted from High-2 and High-3 cultures at the 10th generation of the long-term production experiment (i.e., the earlier high-production state prior to production decline). Following the retransformation, SARS-CoV-2-mCherry fusion expression in the cured strains was found to be restored to levels comparable to those observed at the early production stage (**Figure 4.3**). This suggests that the production decline was primarily caused by changes accumulated in the expression plasmid during extended cultivation, rather than by stable changes in the host genome.

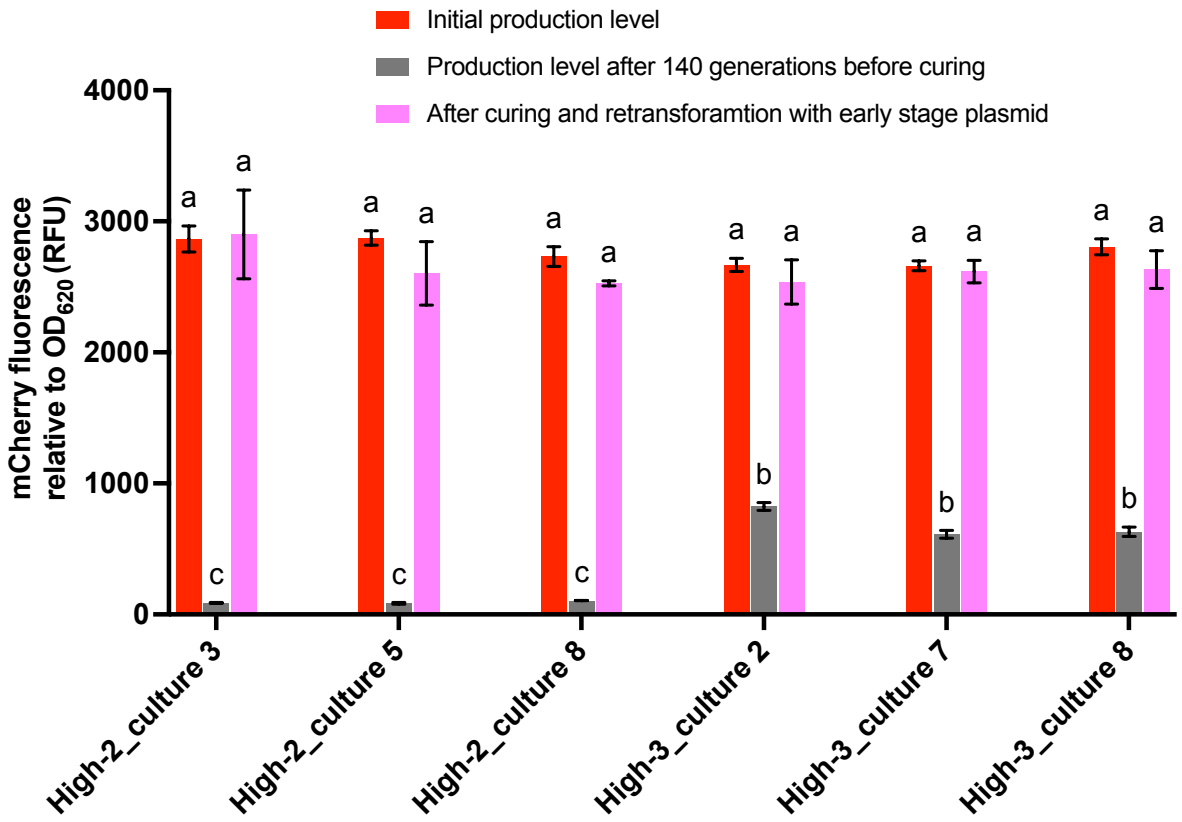


Figure 4.3: SARS-CoV-2 mCherry fusion protein production by selected low-producing colonies from High-2 and High-3 cultures after 140 generations, before and after curing and retransformation with early-stage plasmid. Initial production levels of all eight cultures at the early-stage 10th generation are shown as a control. The selected, cultured colonies (red points in Figure 4.2) were cured of their production plasmid and retransformed with production plasmids extracted from the corresponding cultures at the 10th generation (initial high-production state). Both the selected colonies before curing and the retransformants after curing and retransformation were cultured in 48-well microtiter plates in 500 µL BMMD + CSM – Leu – Met medium for 3 days, before measuring total mCherry fluorescence with normalisation to OD_{620}. The grey bars (before curing) and pink bars (after curing and retransformation) represent means ± SEM of three biological replicates for each

selected colony. Means with different letters are significantly different (two-way ANOVA interaction): $F(10, 32) = 9.648$, $p < 0.0001$.

4.3.4 Whole plasmid sequencing revealed mutations in the production plasmids

To characterize potential changes in the production plasmid that might have contributed to the observed production decline, whole-plasmid sequencing was performed on plasmids isolated from the selected low-expressing colonies (highlighted as red datapoints in Figure 4.2). In high-copy-number plasmid systems, such as the whole-2-micron plasmid system used in this thesis, individual plasmid molecules can accumulate mutations independently within an individual cell over multiple generations of replication and segregation. This can result in intracellular plasmid heterogeneity, particularly under high metabolic burden that imposes selective pressure for reduced production [22; 200]. The six selected low-expressing colonies derived from evolved cultures at the 140th generation (one colony from each High-2 culture 3, 5 and 8, one from each High-3 culture 2, 7 and 8) were therefore prone to harbouring mixed populations of plasmid variants. To address this potential heterogeneity and enable characterization of individual plasmid genotypes, plasmids were first extracted from these six colonies (resulting in six plasmid preparations). These plasmids were then retransformed into the corresponding cured High-2 and High-3 strains derived from the early stage of the long-term production (10th generation). This ensured an isogenic genetic background of the retransformants for each strain, and so the only variable was the plasmid introduced. For each plasmid preparation, two independent retransformants were selected (resulting in six for High-2 and six for High-3), with each

retransformant presumed to carry a single-type plasmid (sp = single-type plasmid). Specifically, High-2_sp1 and sp2 retransformants carried plasmids extracted from the colony derived from High-2 culture 3; High-2_sp3 and sp4 carried plasmids extracted from the colony derived from High-2 culture 5; High-2_sp5 and sp6 carried plasmids extracted from the colony derived from High-2 culture 8; High-3_sp1 and sp2 carried plasmids extracted from the colony derived from High-3_culture 2; High-3_sp3 and sp4 carried plasmids extracted from the colony derived from High-3_culture 7; High-3_sp5 and sp6 carried plasmids extracted from the colony derived from High-3_culture 8. The aim was not to achieve comprehensive coverage of all potential plasmid variants, but rather to assess whether heterogeneity was present within plasmid populations of individual colonies. This strategy provided a sufficient breadth of coverage to achieve this objective within the limited timeframe of the study. These retransformants were cultured in BMMD + CSM – Leu – Met medium for three days, and production levels were measured for each retransformant (**Figure 4.4**). The results showed that for High-2 retransformants, production levels of all selected retransformants were comparable to the empty vector control, that is, the parental plasmid backbone lacking the SARS-CoV-2-mCherry expression cassette (non-productive control). In contrast, the High-3 retransformants displayed more variable expression levels. For example, High-3_sp3 and sp4 retransformants showed clear differences in production despite carrying plasmids originating from the same selected colony. This supports the suggestion that the plasmid populations within each selected colony might carry different mutations. Plasmid DNA was then extracted again from these retransformants to ensure recovery of uniform plasmid types. These purified plasmid samples were subsequently sent to whole-plasmid sequencing for identification of mutations that might explain the reduced production phenotype.

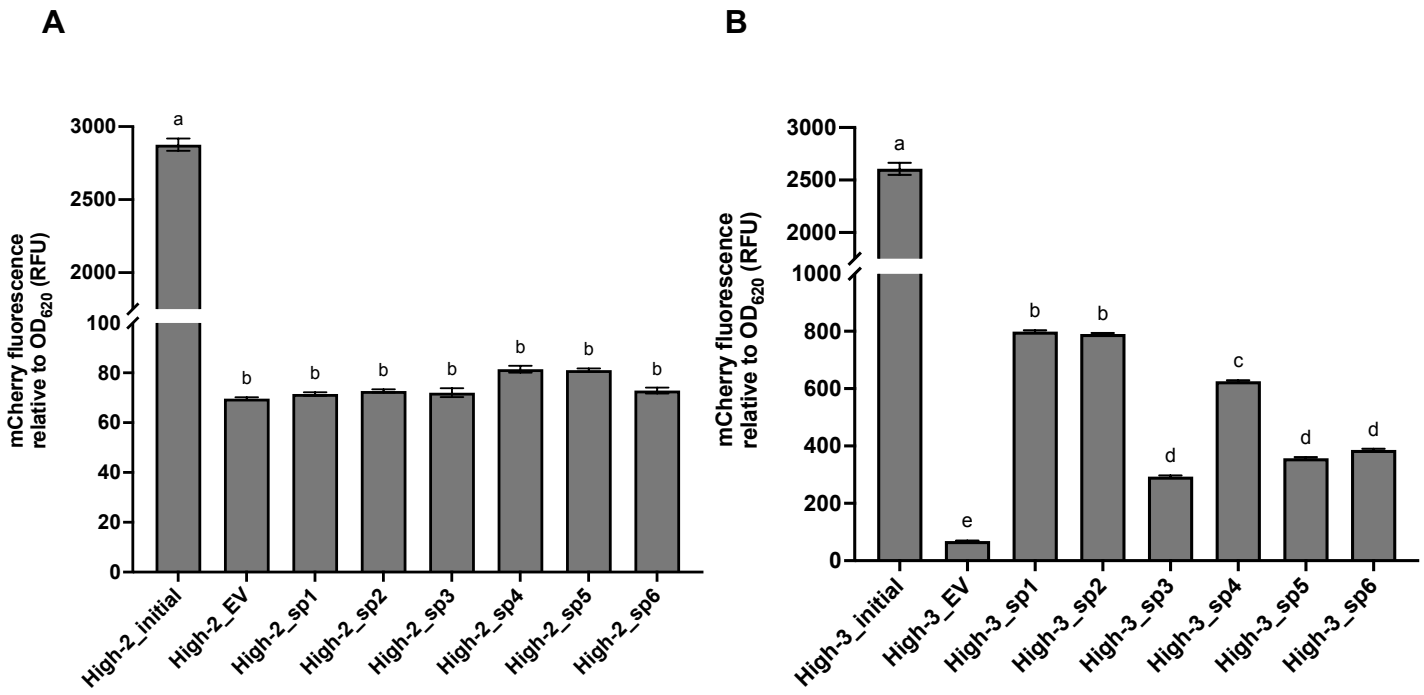


Figure 4.4: SARS-CoV-2-mCherry fusion protein expression from single-type plasmids isolated from low-expressing colonies selected from evolved cultures.

High-2 and High-3 strains at the early production stage (10th generation) cured of plasmids were retransformed with plasmids extracted from selected colonies derived from evolved cultures at 140th generation (red datapoints in **Figure 4.2**). Each retransformant was assumed to have taken up only a single plasmid molecule from the relevant extract. Two independent retransformants were selected for plasmids from each colony (High-2_sp1 and sp2 from High-2 culture 3; High-2_sp3 and sp4 from High-2 culture 5; High-2_sp5 and sp6 from High-2 culture 8; High-3_sp1 and sp2 from High-3_culture 2; High-3_sp3 and sp4 from High-3_culture 7; High-3_sp5 and sp6 from High-3_culture 8; sp = single-type plasmid). The High-2 and High-3 cured strains retransformed with production plasmids extracted at the initial high-production stage of 10th generation (High-2_initial and High-3_initial) and empty vector (EV) lacking the production cassette (High-2_EV and High-3_EV) were used as controls.

Production levels, measured as mCherry fluorescence normalised to OD_{620}, for High-2 (A) and High-3 (B) extracted single-type plasmids expressed in the relevant cured strain backgrounds are presented as mean values ± SEM for three biological replicates. Means with different letters are significantly different (Ordinary one-way ANOVA: (A) $F(7, 16) = 4424$, $p < 0.0001$; (B) $F(7, 16) = 1498$, $p < 0.0001$).

Whole-plasmid sequencing achieved a mean sequencing depth of 15.2x across the plasmid. The sequencing results revealed that all plasmids carried mutations, with some mutations occurring in the same regions but varying in type across the plasmids (**Table 4.1**). Notably, even plasmids derived from the same original colony displayed distinct mutation types, supporting the suggestion that plasmid populations within individual cells can diverge over generations, especially under high burden that can drive adaptive sequence changes [22; 200]. The mutations identified were in diverse functional regions of the plasmid, including *FRT* sites, the SARS-CoV-2 spike protein fragment coding sequence, the *MET17* promoter, the mCherry coding sequence, the FLP recombinase gene, the *LEU2* promoter, the expression cassette terminator region, and non-coding regions. Importantly, all sequenced plasmids contained at least one mutation with the potential to reduce production efficiency. This strongly suggests that production-related mutations, such as those in the *FRT* regions, *MET17* promoter and SARS-CoV-2 spike protein fragment coding sequence, likely arose during the long-term production experiment and were selected as an adaptive response to relieve metabolic burden. It is possible that recombinant protein production itself increased mutation rates, because mutation rates can increase under stress in microbial systems [76; 82; 201; 202]. The possibility that these mutations might have emerged during

later steps such as single-colony isolation, short-term culturing, or the transformation procedure cannot be entirely excluded. However, cells with the reference plasmids from the 10th generation (High-2_initial and High-3_initial) were subjected to the same workflow of colony isolation, plasmid extraction, retransformation, re-extraction and sequencing. The mutations as described in **Table 4.1** were identified by aligning the sequences of evolved plasmids from the 140th generation (High-2_sp1-6 and High-3_sp1-6 sequences) against these 10th-generation references (High-2_initial and High-3_initial sequences). Therefore, it is very unlikely that the observed mutations originated from artefacts introduced during those later steps. Overall, the evidence strongly supports that the mutations identified in evolved plasmids arose during long-term cultivation.

Table 4.1: Nature of mutational changes characterised by sequencing of plasmids extracted from the indicated strains after production for 140 generations (the specific sequence changes are detailed in Table Appendix C.1).

Plasmids from strains after production for 140 generations	Mutation types*	Locations of mutational changes
High-2_sp1 (from the selected colony of Culture 3)	type 1a, type 1c, type 2a, and type 8a	*FRT* regions, SARS-CoV-2 protein seq and noncoding region
High-2_sp2 (from the selected colony of Culture 3)	type 2a	SARS-CoV-2 protein seq
High-2_sp3 (from the selected colony of Culture 5)	type 2b	SARS-CoV-2 protein seq
High-2_sp4 (from the selected colony of Culture 5)	type 1a, type 1c, type 2a, and type 8a	*FRT* regions, SARS-CoV-2 protein seq and noncoding region
High-2_sp5 (from the selected colony of Culture 8)	type 2a	SARS-CoV-2 protein seq
High-2_sp6 (from the selected colony of Culture 8)	type 4 and type 6	mCherry seq and FLP seq.
High-3_sp1 (from the selected colony of Culture 2)	type 1e and type 1f	*FRT* regions
High-3_sp2 (from the selected colony of Culture 2)	type 1b, type 5 and type 8c	*FRT* regions and terminator of expression cassette
High-3_sp3 (from the selected colony of Culture 7)	type 3	p*MET17* region
High-3_sp4 (from the selected colony of Culture 7)	type 1b, type 1d and type 8b	*FRT* regions and noncoding region
High-3_sp5 (from the selected colony of Culture 8)	Type 3, type 7 and type 8a	p*MET17*, LEU2 promoter, noncoding region
High-3_sp6 (from the selected colony of Culture 8)	type 1a, type 1c and type 3	*FRT* regions and p*MET17*

***Mutation types refer to different regions of the plasmid where sequence changes were identified. Type 1: *FRT* regions; Type 2: SARS-CoV-2 spike protein fragment coding sequence; Type 3: *pMET17* promoter region; Type 4: *mCherry* coding sequence; Type 5:**

terminator region of the expression cassette (downstream of mCherry); Type 6: FLP recombinase coding sequence; Type 7: *LEU2* promoter region; Type 8: Non-coding regions. The lower cases a, b, c, d, e and f indicate distinct mutations within that region. Detailed sequence changes corresponding to each mutation type and subtype are provided in Table Appendix C.1.

Analysis of the plasmids from the High-2 retransformants (High-2 sp1-6) showed that all contained mutations within the expression cassette: five with mutations in the SARS-CoV-2 spike protein fragment coding sequence and one with a mutation in the mCherry coding sequence (**Table 4.1**). These mutations likely explain why all High-2 retransformants exhibited very low mCherry levels (**Figure 4.4A**), as each example is predicted to cause either a frame shift in the SARS-CoV-2 spike protein fragment sequence or premature termination of translation. Either scenario is expected to decrease measurable mCherry fluorescence. Notably, High-2_sp1 and High-2_sp4 plasmids shared identical mutations, as did High-2_sp2 and High-2_sp5 plasmids, even though they were derived from different evolved cultures (**Table 4.1**). This suggests that the selective pressures in the High-2 cultures may have favoured the emergence of these specific mutations. For the plasmids from the High-3 retransformants, overall expression levels were also significantly lower than the initial high-production state, but three of them (High-3_sp1, 2 and 4) conferred significantly higher SARS-CoV-2-mCherry expression compared to the other three (High-3_sp3, 5 and 6) (**Figure 4.4B**). Sequencing revealed that the lower-producing plasmids all carried a single nucleotide substitution in the *MET17* promoter region, whereas the relatively higher-producing plasmids all carried mutations in the *FRT* sites (**Table 4.1**). Notably, High-3_sp1 carried mutations only in the *FRT* sites, and High-3_sp3 harboured a mutation exclusively in the *MET17* promoter (**Table 4.1**), implying that mutations at either of these locations alone can be sufficient to reduce production.

Additionally, it was considered that mutated protein products might still be expressed, but with undetectable fluorescence. To exclude this possibility, all High-2 and High-3 retransformants were cultured in BMMD + 2xCSM – Leu – Met medium for three days

and culture supernatants were collected and analysed by SDS-PAGE. The SDS-PAGE results confirmed that the protein expression was indeed reduced compared to the initial high-production and empty vector controls (**Figure 4.5**). The potential impact of production burden on growth was also assessed by measuring the specific growth rates of strains at the early (10th generation) and the late (140th generation) stages of the long-term production experiment, as well as the retransformants carrying single-type plasmids (**Figure 4.6**). Strains that exhibited high production at the initial stage showed reduced growth rates compared to the empty vector controls. In contrast, the evolved low-producing cultures at the 140th generation and the retransformants expressing low-producing plasmids showed increased growth rate relative to the initial stage, which either matched or approached that of the empty vector controls. These observations support the inference that declining production over time alleviated the metabolic burden on the cell.

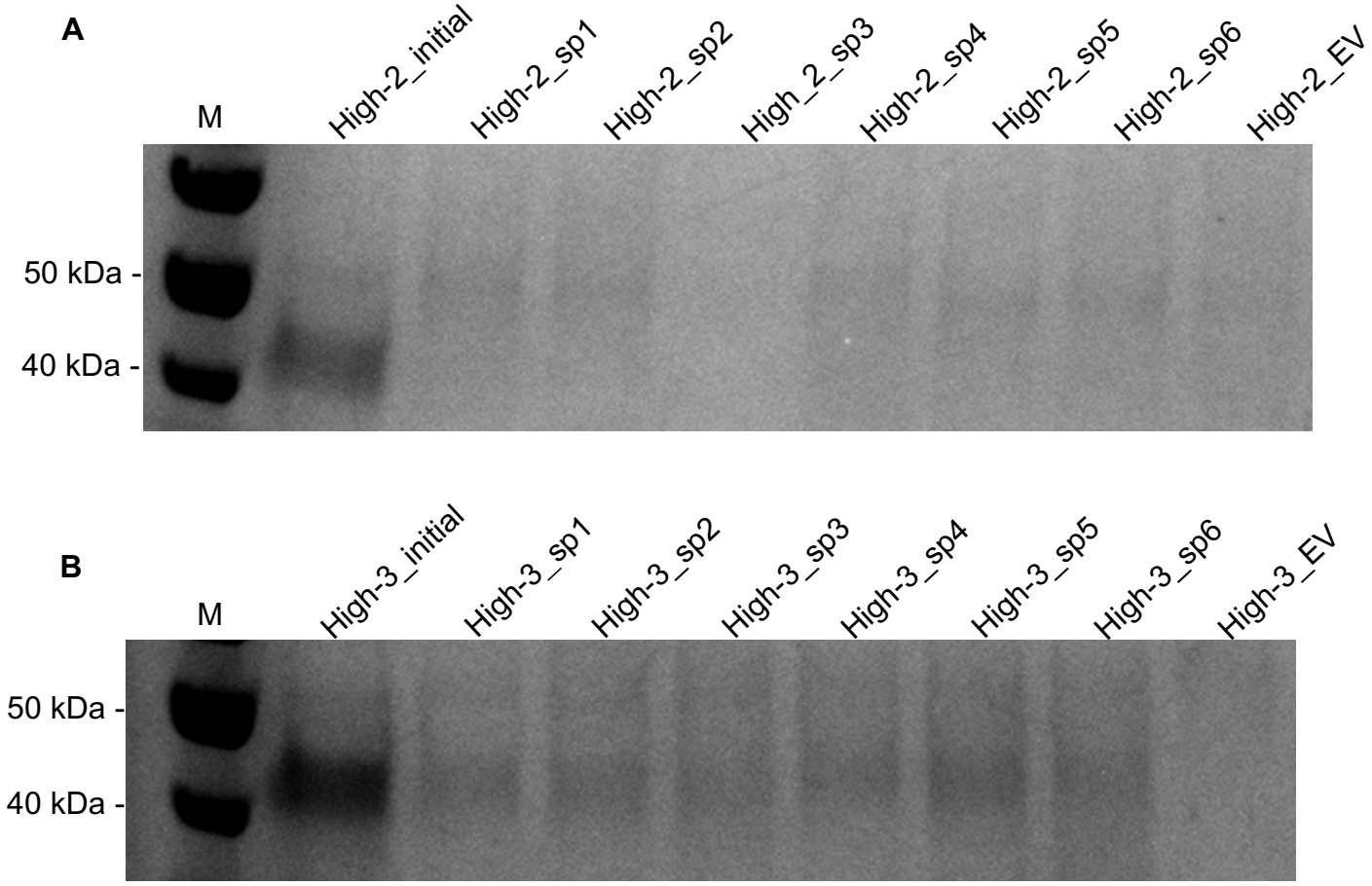


Figure 4.5: SDS-PAGE analysis of SARS-CoV-2-mCherry fusion protein secretion in culture supernatants from retransformants. Supernatants from High-2 and High-3 retransformants shown in **Figure 4.4** were collected. Strains with production plasmids extracted at the initial high-production stage of 10th generation (High-2_initial and High-3_initial) and with the empty vector lacking the production cassette (High-2_EV and High-3_EV) were used as controls. Each retransfromant was cultured in BMMD + 2xCSM – Leu – Met medium for three days, and supernatants were collected for SDS-PAGE analysis. M = molecular weight marker. (A) Secretion of SARS-CoV-2-mCherry fusion protein in supernatants from High-2 retransformants. (B) secretion of SARS-CoV-2-mCherry fusion protein in supernatants from High-3 retransformants.

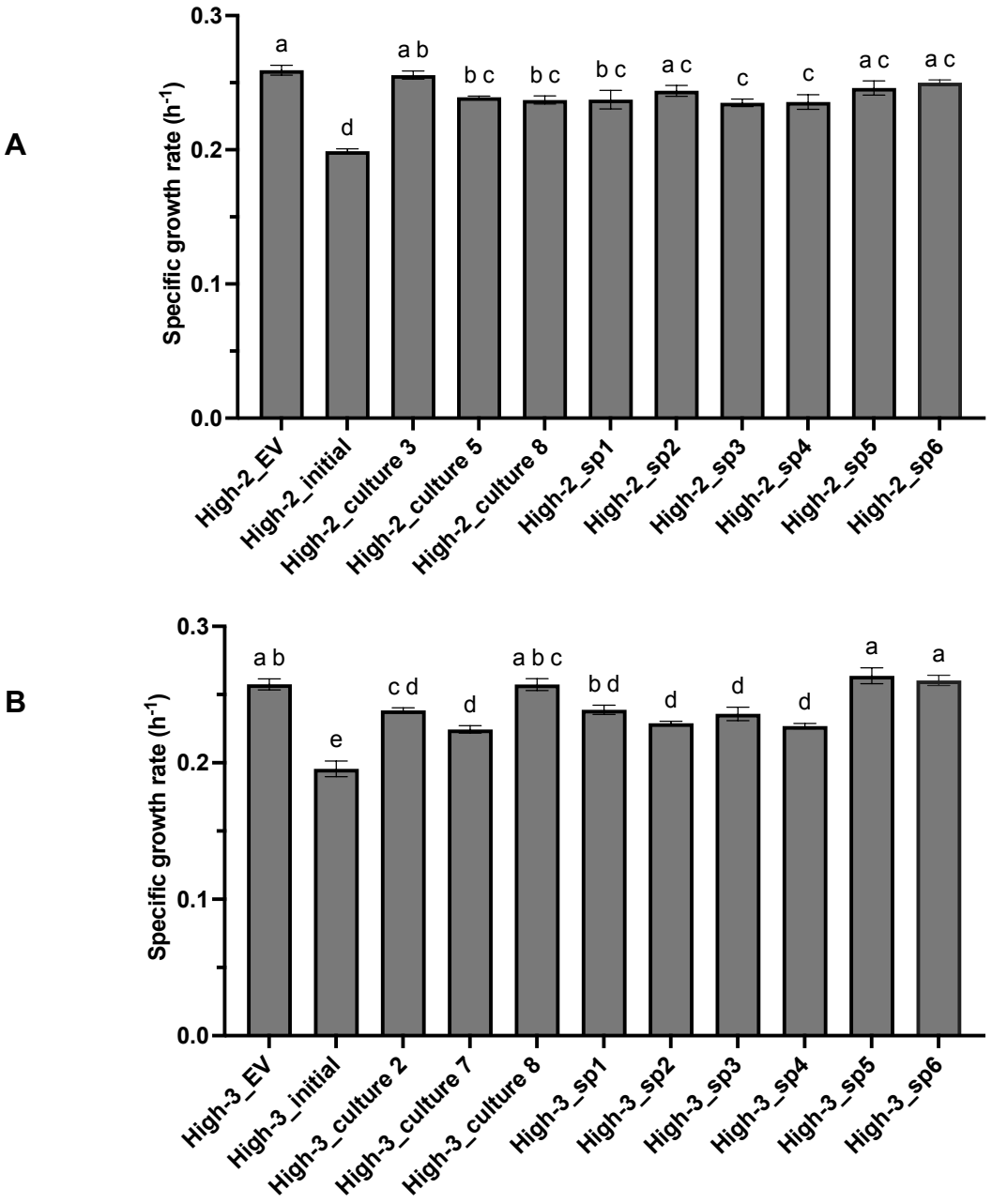


Figure 4.6: Specific growth rates of High-2 and High-3 strains at early and late stages of the production cycle, and retransformants with single-type plasmids derived from the late production cycle. Cells were cultured in sequential bath culture in BMMD + CSM – Leu – Met medium. Specific growth rates were calculated during the exponential growth phase using OD_{620} measurements taken over a 3-hour time interval. Specific growth rates of High-2 (A) and High-3 (B) cultures at the 140th

generation (mix cell and plasmid populations) and the relevant cured strains expressing culture-derived single-type plasmids (High-2_sp1-6 and High-3_sp1-6) are presented as mean values ± SEM for four biological replicates. Empty vector (High-2_EV and High-3_EV) and initial production plasmid (High-2_initial and High-3_initial) were used as controls. Means with different letters are significantly different (Ordinary one-way ANOVA: (A) $F(10, 33) = 16.22$, $p < 0.0001$; (B) $F(10, 33) = 26.99$, $p < 0.0001$).

4.3.5 A single-nucleotide point mutation within the *MET17* promoter can dramatically reduce promoter activity

In yeast-based systems, regulatable promoters are widely employed for the rational design and optimisation of cell factories [203]. In this thesis, the regulatory *MET17* promoter was used to control the expression of the SARS-CoV-2-mCherry fusion protein. As aforementioned, High-3_sp3 harboured a mutation exclusively in the *MET17* promoter. Specifically, the only mutation that had arisen in this plasmid was a single G-to-A point mutation within the CACGTG motif of the *MET17* promoter (**Table Appendix C.1**). The same mutation was also identified in the *MET17* promoter sequences of High-3_sp5 and sp6 based on whole plasmid sequencing results (See **Table 4.1 and Table Appendix C.1**). Retransformants carrying these plasmids (High-3_sp3, sp5 or sp6) exhibited markedly reduced SARS-CoV-2-mCherry expression compared to those carrying High-3_sp1, sp_2 or sp_4 (**Figure 4.4B**). Therefore, it was hypothesised that this single nucleotide substitution might significantly impair *MET17* promoter activity. To test this, the wild-type *MET17* promoter sequence (wt*pMET17*) and mCherry coding sequence were amplified from the production plasmid extracted at the 10th generation (High-3_initial). The mutated *MET17* promoter sequence

(mu*pMET17*) was amplified from the High-3_sp3 plasmid, and the *CYC1* terminator sequence (*tCYC1*) was amplified from the pCM190 plasmid. Reporter plasmids were then constructed using pRS316 as the backbone, incorporating either the wt*pMET17* or mu*pMET17* to generate pRS315-wt*pMET17*-mCherry-*tCYC1* or pRS315-mu*pMET17*-mCherry-*tCYC1*. These reporter plasmids were transformed into the laboratory *S. cerevisiae* strain BY4741. After overnight culture in YNB medium supplemented with methionine, histidine and uracil, mCherry expression was measured to assess promoter activity. The results showed that this single nucleotide substitution reduced *MET17* promoter activity by approximately 88% (**Figure 4.7**), indicating that this mutation alone can significantly reduce *MET17* promoter activity.

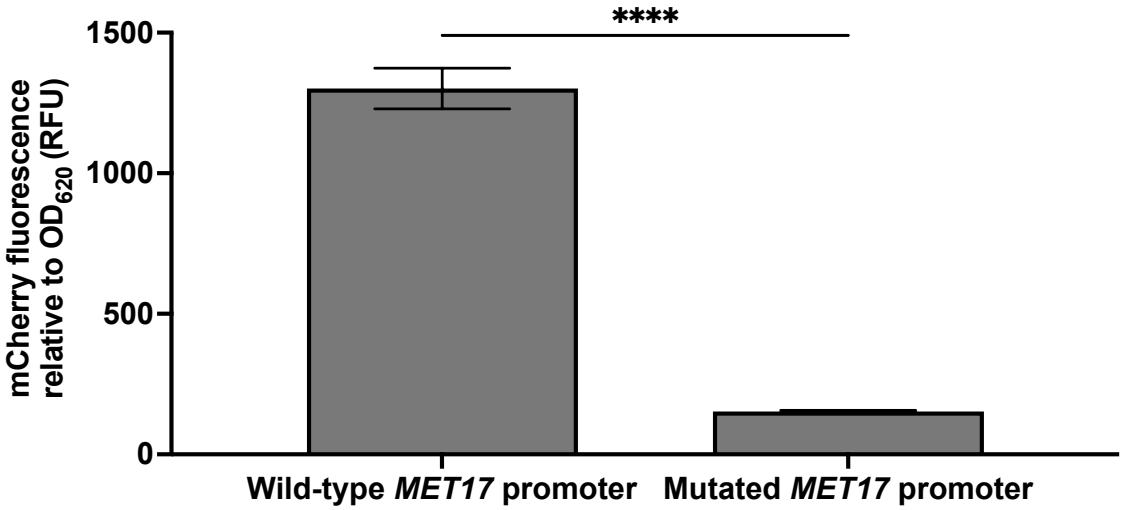


Figure 4.7: Comparison of the wild-type *MET17* promoter activity with the mutated version. The mutated *MET17* promoter contains a single G-to-A point mutation within the CACGTG motif, changing it to CACGTA. Reporter plasmids containing either the wild-type or mutated promoter were used to drive mCherry expression, and promoter activity was assessed by measuring mCherry expression

with normalization to OD_{620}. Values shown are means ± SEM of three biological replicates. Statistical significance (unpaired t-test): $p < 0.0001$ (****).

4.3.6 Plasmid copy number changes in the retransformants carrying single-type plasmids

In addition to promoter activity, another key factor influencing production yield can be the plasmid copy number, because a lower plasmid copy number means fewer copies of the gene of interest [204]. The yeast 2-micron plasmid system relies on a site-specific recombination mechanism mediated by the FLP recombinase and its target sites (*FRTs*) to maintain high plasmid copy number [205]. Mutations in the *FRT* sites can therefore disrupt this amplification system, potentially contributing to the production decline described here. Given that High-3_sp1 carried mutations exclusively in the *FRT* sites (**Table 4.1**) and showed an approximate 68% reduction in production compared to the High-3_initial control (**Figure 4.4B**), it was hypothesized that mutations identified in the *FRT* sites of plasmids might affect plasmid copy number and, therefore, also the production level. To test this, droplet digital PCR (ddPCR) was performed to quantify plasmid copy number per genome in those High-2 and High-3 retransformants carrying the relevant single-type mutated plasmids (High-2_sp1-6 and High-3_sp1-6; see **Table 4.1**). Plasmid copy number per genome was calculated as the ratio of plasmid DNA to genomic DNA concentrations (**Figure 4.8**).

For the High-2 retransformants, the empty vector control (High-2_EV) exhibited the highest plasmid copy number, while the initial high-production plasmid from the 10th generation (High-2_initial) had the lowest (**Figure 4.8A**). Although High-2_sp1 and

High-2_sp4 contained mutations in the *FRT* sites, all High-2 retransformants carrying mutated plasmids (High-2_sp1-6) displayed significantly higher copy numbers than High-2_initial, albeit still lower than High-2_EV.

In contrast, the High-3 retransformants suggested a clearer link between production level and plasmid copy number. Among the High-3 retransformants, higher-producing strains (High-3_sp1, sp2 and sp4) were associated with lower plasmid copy numbers, while the poorer producers (High-3_sp3, sp5 and sp6) exhibited significantly higher plasmid copy numbers (**Figure 4.4B and Figure 4.8B**). Importantly, those high-production plasmid variants shared similar mutations in the 8-bp spacer region of the *FRT* sites, and those mutations appeared to be the only functional mutations potentially affecting production, as other identified mutations in those high-production plasmid variants were in non-coding regions or the terminator region. By contrast, the lower-production variants all carried the same single G-to-A mutation in the *MET17* promoter that reduces promoter activity by ~ 88% as shown in **Figure 4.7**. A minimal *FRT* site consists of 13-bp arms flanking an 8-bp spacer (**Figure Appendix D.1**). The spacer region is critical for strand pairing, crossover and cleavage during FLP-mediated amplification in the 2-micron plasmid system [206]. The results suggest the 8-bp spacer region may have been a mutational hotspot in the 2-micron plasmid system used in this study, as all identified mutations in the *FRT* sites in both High-2 and High-3 plasmid variants altered the spacer sequence, though the exact changes varied among variants (**Table Appendix C.1**). Despite these *FRT* mutations, the High-3 retransformants with lower expression levels (High-3_sp3, sp5 and sp6) tend to have higher plasmid copy number than the retransformants with higher expression levels (High-3_sp1, sp2 and sp4). This pattern is similar to the observation in the High-2

retransformants, where the high-producing strain (High-2_initial) had fewer plasmid copies than the low-producing strains (High-2_sp1-sp6) (**Figure 4.4A and 4.8A**). However, in contrast to High-2, the High-3_initial control combined both higher production and higher plasmid copy number compared to the High-3_sp1, sp2 and sp4 strains (**Figure 4.4B and Figure 4.8B**). Since High-3_sp1, sp2 and sp4 carried functional mutations only in the *FRT* sites, these results suggest that plasmid copy number is shaped not only by expression burden but also by sequence changes at the *FRT* sites. It was therefore questioned whether these mutated *FRT* sites impaired FLP/*FRT*-mediated recombination, potentially explaining the altered copy number patterns.

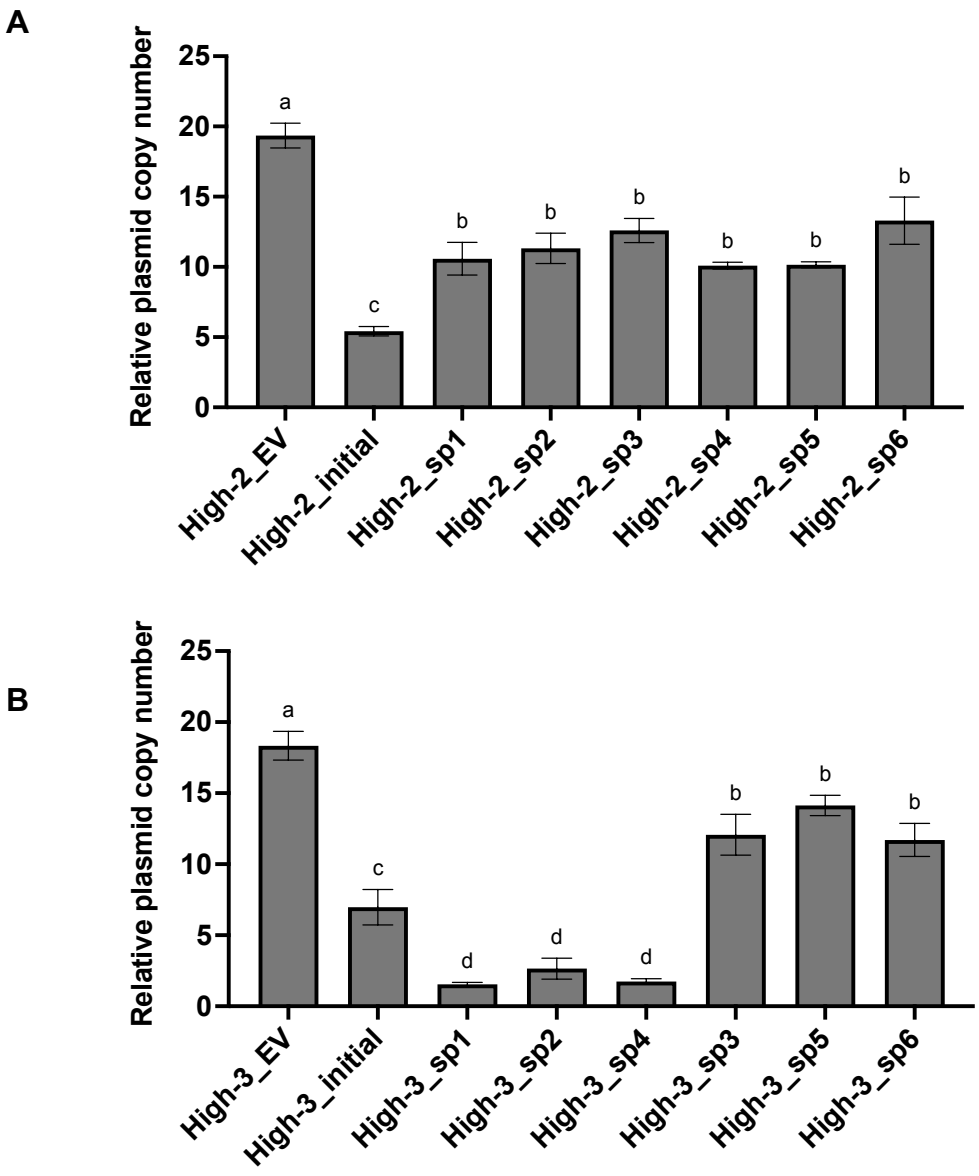


Figure 4.8: Quantification of production plasmid copy number per genome in High-2 and High-3 retransformants. Empty vector (EV) and initial high-production plasmid from the 10th generation were used as controls. Plasmid copy number per genome was quantified in High-2 (A) and High-3 (B) retransformants carrying the relevant mutated plasmids High-2_sp1-6 and High-3_sp1-6 (as listed in **Table 4.1**) using ddPCR. The *ACT1* gene was used as a reference gene for genomic DNA concentration. Plasmid copy number per genome was calculated as the ratio of plasmid concentration (copies/µL) relative to genomic DNA concentration (copies/µL).

Data represent mean values ± SEM of three biological replicates. Means with different letters are significantly different (Ordinary one-way ANOVA: (A) $F(7, 16) = 17.40$, $p < 0.0001$; (B) $F(7, 16) = 45.32$, $p < 0.0001$).

4.3.7 *FRT/*FLP-mediated recombination efficiency decreased with the mutated *FRT* sites

To test whether the observed changes in plasmid copy number were linked to impaired *FRT/*FLP-mediated recombination, an assay was designed using plasmid constructs containing a kanamycin resistance gene (*KanMX*) flanked by wild-type or mutated minimal *FRT* sequences in the same orientation (**Figure 4.9A**). All four types of mutated *FRT* site derived from High-3_sp1, sp2, sp4 and sp6 plasmids were tested, using a construct with wild-type *FRT* sites as a control. These particular plasmids were chosen for analysis because the mutations identified in the High-3_sp1, sp2 and sp4 plasmids, which supported relatively high production levels yet lower plasmid copy number than others (High-3_initial, High-3_sp3, sp5 and sp6) (**Figure 4.4B**), were restricted to the *FRT* sites. Specifically, High-3_sp1 and sp4 each contained different mutation subtypes in both *FRT* sites, while High-3_sp2 had mutation in only one *FRT* site (see **Table 4.1 and Table Appendix C.1**). High-3_sp6 was also included, as it contained mutations in both *FRT* sites along with a mutation in the *MET17* promoter, showing a higher plasmid copy number but a reduced production level (see **Figure 4.4B** and **Figure 4.8B**).

All plasmid constructs, containing either wild-type or the mutated *FRT* sites, were transformed into a strain of *S. cerevisiae* BY4741 that had been engineered to

overexpress FLP recombinase under the control of the strong, galactose-inducible *GAL1* promoter (see **Section 4.2.8**). Upon induction with galactose, active FLP should catalyse recombination between functional *FRT* sites, resulting in excision of the *KanMX* cassette. Successful excision should render transformants sensitive to kanamycin, whereas failure to recombine would result in retention of the cassette and continued resistance, indicating impaired recombination (**Figure 4.9A**). As expected, the control construct with wild-type *FRT* sites showed high recombination efficiency (**Figure 4.9B**), i.e. successful excision of the *KanMX* cassette. In contrast, all four mutated *FRT* variants exhibited significantly reduced recombination efficiencies, ranging from 3% to 21.5% of the control level, indicating that these mutations substantially impaired FLP recognition or recombination activity. Interestingly, High-3_sp6 displayed higher plasmid copy number (**Figure 4.8B**) but much lower recombination efficiency (**Figure 4.9B**). This suggests that while the *FRT* mutations reduced the efficiency of the recombination-based plasmid amplification mechanism, plasmid copy number was not solely determined by this mechanism. In the case of High-3_sp6, the variant also contained the mutation in the *MET17* promoter, which substantially impaired promoter activity and reduced production level to that comparable with the empty vector control (**Figure 4.7**). These findings suggest that length-altering mutations within the spacer region disrupt *FRT*/FLP-mediated recombination, which can lead to reduced plasmid copy number and, consequently, lower recombinant protein expression, except when the functional mutations are not restricted to the *FRT* sites.

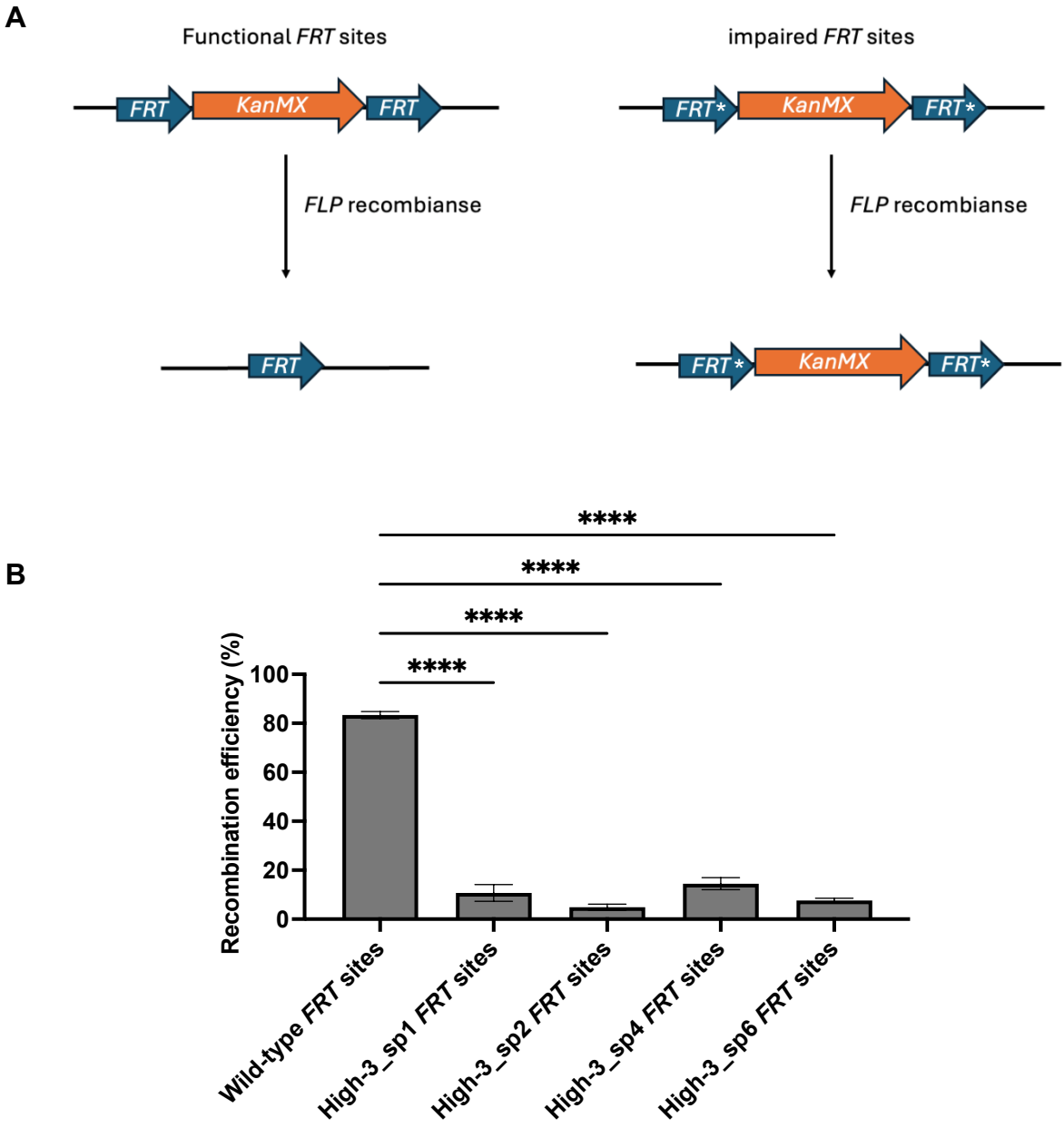


Figure 4.9: Recombination efficiency of FLP recombinase with wild-type and mutated *FRT* sites. (A) Schematic of the FLP-mediated recombination efficiency assay design. Plasmids were constructed with a *KanMX* resistance gene flanked by either functional (left) or mutated (right) *FRT* sites arranged in the same direction. Upon expression of FLP recombinase, recombination between functional *FRT* sites excises the *KanMX* cassette, leading to loss of kanamycin resistance; plasmids with impaired *FRT* sites fail to support FLP-mediated recombination, and the *KanMX*

cassette is retained, maintaining kanamycin resistance. (B) Experimental measurement of FLP-mediated recombination efficiency for constructs containing wild-type or mutated FRT sites derived from High-3_sp1, sp2, sp3 and sp6 plasmids. *S. cerevisiae* BY4741 engineered to express FLP recombinase under the control of the *GAL1* promoter were transformed with these plasmid constructs. Transformants were precultured in YNB medium with 2% glucose, supplemented with uracil and methionine to maintain selection. For induction, cultures were serially passaged in selective YNB medium with 2% galactose as the carbon source for approximately 50 generations to ensure continuous FLP expression. After induction, cultures were diluted and plated onto non-selective YPD agar and YPD agar containing G418. Recombination efficiency was defined as the proportion of colonies sensitive to G418 relative to the total number of colonies on YPD. Data represent mean values ± SEM of three biological replicates. Statistical significance (Ordinary one-way ANOVA): $F(4, 10) = 247.3$, $p < 0.0001$ (****).

4.3.8 Natural variation at the *MMS21* locus shows no significant association with production stability but suggests a trend

Previous studies have demonstrated that many host factors can influence 2-micron plasmid stability in *S. cerevisiae* [180; 207-209]. A notable example is natural variation in the *MMS21* gene, as described in **Section 4.1** of this chapter. Briefly, a single amino acid change in *MMS21* has been shown to impact plasmid stability: the ancestral Thr69 allele is associated with greater 2-micron plasmid retention, while the more recently derived Ile69 allele is linked to increased 2-micron instability [180]. Based on these findings, it was hypothesised that variation at the *MMS21* locus might contribute

to the production decline observed over the long-term production experiment in this study. In particular, strains maintaining higher production stability were expected to be more likely to carry the Thr69 allele, whereas strains showing greater production decline over 140 generations might tend to carry the Ile69 allele. To test this, 30 haploid strains selected from the library provided by Phenotypeca (as shown in **Figure 3.6A**) were analysed. Colony PCR was performed to amplify the *MMS21* locus, and the cleaned PCR products were sequenced to identify the presence of either the Thr69 allele or the Ile69 allele. Among the 30 strains, 18 carried the Thr69 allele and 12 carried the Ile69 allele. Comparison of production stability between the two genotype groups revealed no statistically significant difference (**Figure 4.10**).

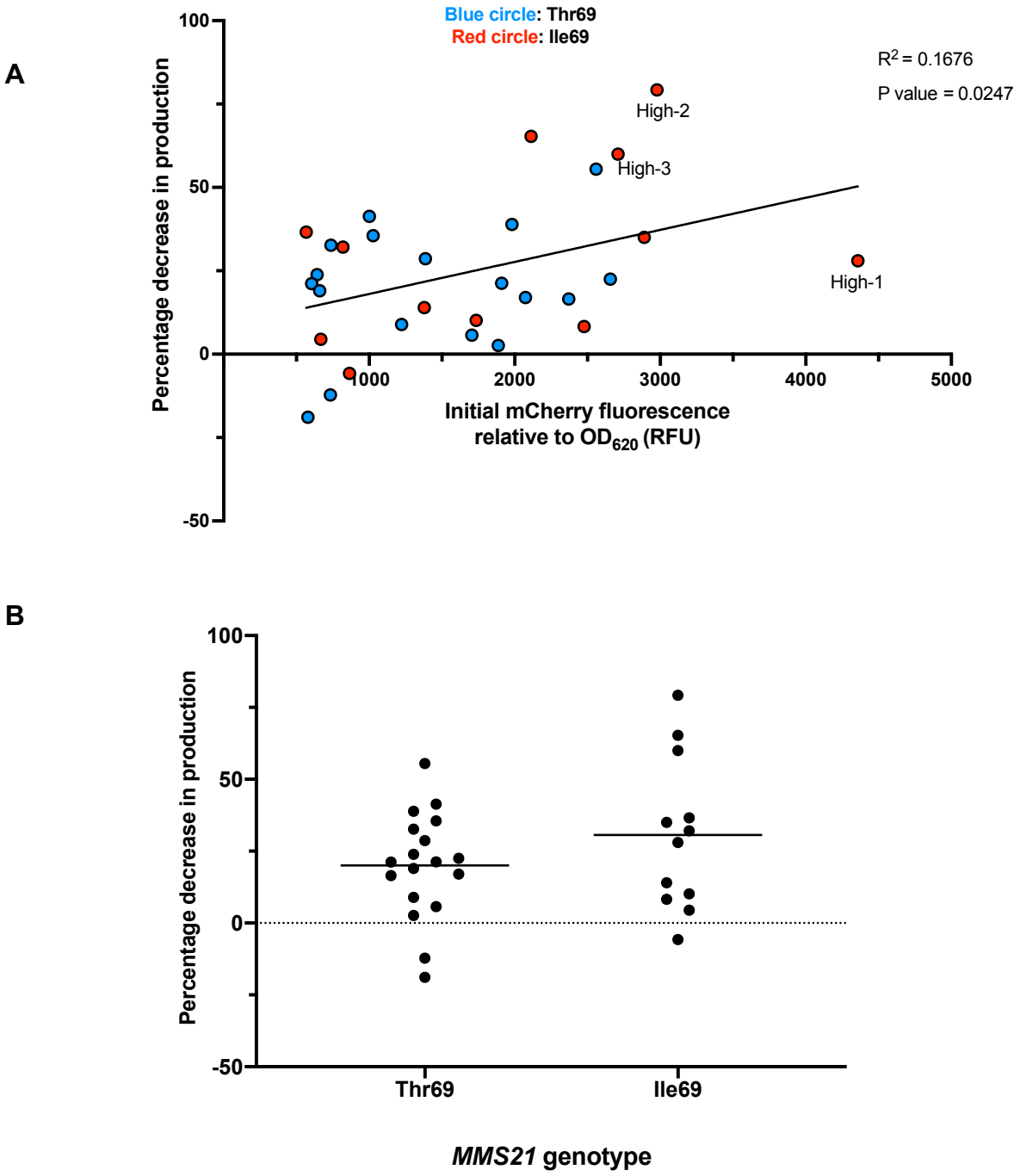


Figure 4.10: Relationship between *MMS21* genotype and long-term production stability across 30 selected haploid *S. cerevisiae* strains. Colony PCR was used to amplify the *MMS21* locus in the 30 haploid strains described in **Figure 3.6A,** and the cleaned PCR products were sequenced to identify the presence of either the Thr69 or Ile69 allele. (A) Including data derived from **Figure 3.6A**, the relationship among initial production levels, percentage decrease in production after 140 generations, and

MMS21 genotype for each strain. Blue circles represent strains with the Thr69 allele, and red circles represent those with the Ile69 allele. (B) Comparison of the percentage decrease in production over 140 generations between the two genotype groups. The 30 strains were grouped by their *MMS21* allele, with lines indicating the mean percentage decrease within each group.

4.4 Discussion

This chapter mainly investigated the genetic factors that contributed to production instability in high-producing yeast strains carrying 2-micron-based expression plasmids. The results showed that multiple types of mutations, including those within the heterologous protein coding sequence, the promoter regions, and the *FRT* sites, were repeatedly associated with declines in productivity. These findings suggested that 2-micron-plasmid-based systems were vulnerable to mutational events that undermined long-term stability, and that both plasmid architecture and host background played important roles in shaping the trajectory of production decline.

4.4.1 The complexity of production decline during extended cultivation

The observation that recombinant protein production could be partially restored in some High-2 and High-3 evolved cultures after a repression phase suggests that non-genotypic factors, such as phenotypic switching or cellular adjustments that alleviate the resource burden of recombinant protein expression, contributed to the decline in recombinant protein yield. This supports the hypothesis that production burden can lead to a shift in population structure through the selection of pre-existing phenotypic variants with lower expression, which in turn improves overall cellular fitness. Similar production burden-driven shifts in population structure, favouring pre-existing low-burden variants, have been reported in microbial systems engineered for overproduction of metabolites and proteins [77; 80]. These findings highlight the complexity of production decline during long-term cultivation, where both non-genetic phenotypic heterogeneity and genetic diversification can influence the stability of recombinant expression. Previous studies have shown that stress adaptation in yeast populations often involves a combination of phenotypic heterogeneity and genetic

diversification, enabling subpopulations to explore different adaptive strategies under pressure [210; 211]. In the context of this thesis, heterologous protein production itself represents a chronic stressor that can drive similar responses. Understanding the interplay between genotypic and non-genotypic heterogeneity in production systems is essential for improving strain stability and performance. Irreversible genetic mutations can potentially be managed by constructing strains with lower mutation rates. This might be achieved by deleting unstable genomic elements and enhancing DNA repair pathways [212]. However, reversible phenotypic heterogeneity is more difficult to control, as it can arise from both stochastic and deterministic factors [213-216]. Stochastic variation results from random fluctuations in molecular processes, such as variability in gene expression due to promoter noise or uneven distribution of transcription factors. For example, cell-to-cell differences in the basal activity of stress-responsive promoters like *pHSP12* can generate variable stress tolerance among genetically identical cells [108]. On the other hand, deterministic factors relate to predictable sources of variation, such as cell cycle stage and replicative age. A recent study has shown that introns in yeast can act as regulatory elements that mediate inducible phenotypic heterogeneity in gene expression within an isogenic population [216]. To reduce the impact of such heterogeneity, genetic circuits that help stabilize or enhance biosynthesis during prolonged cultivation have been developed [80; 145]. The non-genotypic aspect of the production decline is investigated further in Chapter 5.

Regarding genotypic changes, whole plasmid sequencing revealed that even though only one colony was selected from each of the three lowest-expressing cultures at the 140th generation of the long-term production experiment, and only two retransformants

were analysed per colony, sequenced plasmids displayed distinct mutation patterns. The mutations identified were different in their locations, types and subtypes, demonstrating that functionally convergent outcomes can arise through diverse mutational trajectories. This is consistent with previous studies in which populations evolved toward similar phenotypic outcomes through genetically distinct paths [217; 218]. It suggests that under production stress, independent cultures can explore multiple evolutionary solutions to alleviate metabolic burden, leading to genetically diverse, low-producing variants. Furthermore, the retransformation and sequencing data showed that even plasmids extracted from a single colony isolated from the 140th generation could contain a mixed plasmid population, indicating the presence of intracellular plasmid heterogeneity. Such within-cell variation in plasmid genotypes likely emerges during extended cultivation and represents an underappreciated source of evolutionary diversification in multi-copy plasmid-based expression systems. It is important to note that all mutations likely arose randomly during replication, and that production burden did not generate these events but instead provided selective pressure that favoured certain variants with improved fitness. To clarify the extent to which selection underlies the observed outcomes, a useful control could involve cultivating strains under non-inducing conditions (in BMMD + CSM – Leu + Met to repress expression) for 140 generations, followed by colony isolation, plasmid extraction, retransformation, plasmid re-extraction and sequencing. This would help determine whether the observed mutations are specifically favoured by production-associated selective pressures.

4.4.2 Recurrently selected mutations in 2-micron-plasmid expression systems

4.4.2.1 Instability in the heterologous protein coding sequence

One of the regions most frequently carrying mutations among the sequenced plasmids was the SARS-CoV-2-mCherry coding sequence, particularly in evolved plasmids from High-2 strain. Previous studies have shown that heterologous coding sequences expressed in non-native hosts are often prone to mutations due to factors such as suboptimal codon usage, protein misfolding and associated cellular stress [219-221]. While the present results do not directly demonstrate these mechanisms, they are consistent with the idea that coding regions of heterologous proteins can become recurrent targets of selection, as mutations that reduce or disrupt expression may relieve cellular burden and therefore persist. Interestingly, no mutation was detected in the SARS-CoV-2-mCherry coding sequence in any of the sequenced plasmids from High-3 strain. This difference versus the High-2 strain could suggest different levels of stress tolerance between the two host backgrounds. At the endpoint of the long-term cultivation experiment, High-3 cultures still exhibited higher production levels overall than High-2 cultures, suggesting that the High-3 strain was more resistant to the production burden, with genetic adaptations potentially accumulating more slowly or following a different path.

4.4.2.2 Mutations in the promoter region

Mutations were also frequently identified in the *MET17* promoter driving SARS-CoV-2-mCherry expression, especially in evolved plasmids from the High-3 strain. In yeast-based expression systems, inducible and regulatory promoters are widely employed for the rational design and optimisation of cell factories [203]. In this chapter, the

regulatory *MET17* promoter was used to control the expression of the SARS-CoV-2-mCherry fusion protein. Promoter activity assays have revealed that the CACGTG motif, recognised by the yeast centromere binding factor I (Cbf1), a basic transcription factor involved in the regulation of methionine biosynthesis genes [222], is essential for the full activity of the *MET17* promoter. This observation is consistent with previous research demonstrating that Cbf1 is required for transcriptional activation of the *MET17* gene [223]. Mutations in the core motif or regulatory elements are likely to compromise promoter function and reduce gene expression. Given that reduced expression can alleviate metabolic burden, such mutations can confer a selective advantage under sustained production pressure. The repeated enrichment of promoter mutations in evolved plasmids highlights that promoters can represent frequent points of vulnerability under selective pressure. This emphasises the importance of promoter robustness as a design consideration for long-term biomanufacturing applications. Recent efforts in synthetic biology have focused on engineering more stable and resilient promoters for microbial systems. Strategies include modifying promoter architecture to reduce susceptibility to mutation, insulating core motifs or developing synthetic promoters with increased resistance to sequence drift and stress-induced mutagenesis [224-227]. Such improvements may help extend the functional lifetime of expression systems in industrial processes and reduce the likelihood of production decline due to promoter inactivation. Notably, no promoter mutations were detected in any of the sequenced plasmids from the High-2 strain. One possible explanation is that High-2 plasmids had already accumulated mutations in the heterologous protein coding sequence, which significantly reduced production to the level of an empty vector, and thereby conferred a strong growth advantage. These mutants might have quickly dominated the population, reducing the likelihood of further

enrichment of promoter-level changes. In contrast, the High-3 strain may have evolved more slowly, may have been more resistant to mutation within the coding sequence, or the difference may simply reflect stochastic variation in where mutations first arose. As a result, promoter mutation may have emerged earlier in High-3 cultures as a relatively accessible route to lowering production burden. Further support for this interpretation comes from the observation that three of the sequenced High-3 evolved plasmids (High-3_sp1, sp2 and sp4) carried functional mutations only in the *FRT* region and not in the promoter. These variants also maintained relatively high production levels. This suggests that those High-3 cultures were still in the early or intermediate stages of adaptive decline, and that mutations in the promoter or coding sequence could potentially arise at later time points if cultivation were extended beyond 140 generations.

4.4.2.3 FRT spacer mutations

Mutations were also frequently observed within the *FRT* sites of the 2-micron-plasmid expression system. Specifically, all mutations detected in this region were located within the 8-bp spacer sequence of the *FRT* sites. This spacer is known to be essential for amplification and stability in the 2-micron system [206]. Although previous studies have shown that certain base substitutions or even single-nucleotide insertions or deletions within the spacer region can still permit functional *FRT*/FLP-mediated recombination [228-230], the results presented in this chapter indicate that the recombination efficiency is severely compromised when the spacer sequence is more extensively altered. All tested *FRT* mutants (High-3_sp1, sp2, sp4 and sp6) contained non-identical *FRT* sites, each with at least one spacer region exceeding 10 bp in length. Previous work has indicated that functional FLP recombination is generally limited to

FRT sites with altered spacer lengths of less than 10 bp [228]. Therefore, it can be inferred that bigger alterations in spacer length, particularly beyond the 10 bp threshold, can dramatically impair the FLP*/FRT* recombination system. *FRT* site integrity is evidently important in 2-micron expression systems. Future studies could explore the evolution or engineering of FLP recombinase variants capable of recognizing both wild-type and mutated *FRT* sites, thereby improving tolerance to sequence drift and expanding recombination compatibility [231].

4.4.3 Multiple interacting factors influencing plasmid copy number

Although mutations in the *FRT* sites can impair FLP-mediated amplification and thereby reduce plasmid copy number, this mechanism does not solely determine copy number outcomes. For example, High-2_sp2, sp3 and sp5 evolved plasmids, which only carried mutations within the heterologous protein coding sequence and exhibited production levels comparable to the empty vector control, were found to maintain relatively high plasmid copy numbers compared to High-2_initial. Previous work has shown that the 2-micron plasmid is actively transcribed, producing transcripts that encode essential stability functions including *FLP, REP1, REP2* and *RAF1* [232]. Strong transcription from recombinant expression cassettes can interfere with those 2-micron-encoded transcripts, thereby reducing the expression of those maintenance factors. This provides a likely explanation for why strains with lower expression burden can sustain higher plasmid copy numbers. A similar trend was also observed in High-3_sp3 and sp5 evolved plasmids, which only carried a functional mutation in the *MET17* promoter, showing relatively lower production level but higher plasmid copy number compared to High-3_initial. This contrasts with the general assumption that high plasmid copy number is correlated with high expression. In this chapter, the

inverse was observed more than once: higher plasmid copy number coincided with reduced expression, likely because lower production burden reduces transcriptional interference on 2-micron plasmid functions that might otherwise constrain plasmid maintenance [232]. This interpretation is consistent with previous work indicating an inverse relationship between metabolic burden and plasmid copy number [85]. Additionally, another study has suggested that plasmid copy number decreases as the strength of promoter used increases [233], implying that higher production leads to lower plasmid copy number. In addition to *FRT* integrity and production burden, previous studies have shown that plasmid copy number can be influenced by other variables such as selection marker, origin of replication, strain ploidy, promoter strength and host genetic background [200; 233; 234]. Altogether, these findings highlight the multifactorial nature of plasmid copy number control and its dynamic interplay with plasmid stability and production burden.

4.4.4 Strain background effects on plasmid maintenance

Although no statistically significant difference in production stability was observed between the two *MMS21* genotype groups, the data suggest a potential trend: strains carrying the Thr69 allele appeared to show more stable production, while those with the Ile69 allele tended to have a higher mean percentage decrease in production (**Figure 4.8**). Such a trend would be consistent with previous evidence linking the Ile69 allele to reduced 2-micron plasmid retention [180]. The lack of statistical significance in this analysis, if misleading, might reflect the limited sample size. Only 30 haploid strains were analysed in this chapter, which might not provide sufficient power to detect robust genotype-phenotype associations. In contrast, the previous work demonstrating a significant impact of *MMS21* variation on plasmid retention analysed

more than 1000 *S. cerevisiae* isolates [180], suggesting that larger sample sizes may be necessary to detect robust effects. Moreover, plasmid stability is a multifactorial trait influenced by multiple host factors [180; 207-209]. While *MMS21* allele variation may play a role, its effects could be masked or modulated by other genetic variation. Therefore, while these results do not confirm a significant role for *MMS21* variation in production stability in this experiment, they highlight the potential value of larger-scale analyses to better understand a potential contribution of host genetic background to long-term production performance.

4.4.5 Conclusions and future work

This chapter investigated the causes of production decline in selected high-producing *S. cerevisiae* haploid strains expressing a secreted SARS-CoV-2-mCherry fusion protein over 140 generations (High-2 and High-3). Through plasmid curing and retransformation assays, whole plasmid sequencing, promoter activity assay, plasmid copy number quantification and *FRT*/FLP recombination efficiency tests, the results demonstrated that a primary driver of production instability was the accumulation of mutations within the production plasmid.

It is important to recognise that plasmid stability is influenced by multiple host genetic factors that can modulate plasmid replication, partitioning, amplification and cellular stress responses. The genetic background of the host strain can play a role in determining long-term production stability. As shown in Chapter 3, natural diversity among yeast strains already revealed marked differences in their ability to sustain heterologous protein production. Building on that, future work could extend this approach to a larger panel of genetically diverse strains. In such a context, quantitative

trait locus (QTL) analysis could help to associate production stability with specific host loci. However, this would not bypass the need to understand both genotypic and non-genotypic factors. Mechanistic insights remain essential for rational improvements, such as designing more stable synthetic promoters or engineering expression cassettes with reduced instability. QTL analysis would therefore act as a complementary strategy, reinforcing these approaches by incorporating natural variation at loci associated with stability. Together, such mechanistic and mapping-based strategies could guide the selection or engineering of yeast strains with improved tolerance to production-associated stress and more robust plasmid maintenance, ultimately supporting the development of superior industrial yeast cell factories.

Given the diversity observed even with limited sampling, it is plausible that a much broader spectrum of mutational heterogeneity exists within each evolved culture. Systematic selection of a larger number of colonies and sequencing of more plasmids would likely uncover additional mutation types and might help identify dominant, recurrent mutations contributing to the observed production decline. Such work will be essential for providing a more complete understanding of the evolutionary dynamics undermining bioproduction yields, and for designing strategies to develop more robust yeast expression systems capable of sustained high-yield production in industrial biomanufacturing processes.

5 PHENOTYPIC DIVERSITY IN HIGH-PRODUCING YEAST STRAINS

5.1 Introduction

As described in the **Section 1.3.2**, genetically identical cells can display marked differences in growth rate, gene expression, stress tolerance and metabolic activity, a phenomenon termed phenotypic heterogeneity. This variation arises from various sources, including gene expression noise, epigenetic gene regulation, cell cycle-dependent gene expression or possibly microenvironmental variation within a culture if that is not fully homogeneous [22; 96; 123].

Phenotypic heterogeneity in microbial populations is increasingly considered as an adaptive strategy that enables microbial populations to survive and persist in non-stable or fluctuating environments [211; 214; 235]. By maintaining a range of phenotypes, populations can 'hedge their bets', ensuring that at least some cells are well-suited to new or adverse conditions. For example, housekeeping proteins, which are essential for basal cellular function, generally display low expression noise to maintain homeostasis, whereas stress-response proteins often exhibit high expression variability, enabling rapid and flexible responses to environmental changes [235]. This distribution of expression noise appears to be functionally regulated, ensuring stability in essential processes while allowing variability where it may improve survival under stress. In the context of industrial biomanufacturing, high-level recombinant protein production in microbial hosts imposes a significant metabolic burden on cells. This burden can disrupt cellular homeostasis and induce stress responses, which in turn may reduce growth rates, impair protein synthesis, and

contribute to genetic instability [28; 154; 155; 220]. Although industrial processes aim to maintain stable, high-yield production across all cells in a culture, phenotypic heterogeneity, such as that arising from stochastic fluctuations in gene expression, can lead to variation in production levels between individual cells. Subpopulations with lower production levels and burden may grow more rapidly, having an advantage during cultivation. Over time, these low-producing variants can become enriched within the population, ultimately resulting in an overall decline in production efficiency.

Chapter 4 of this thesis investigated the evolutionary dynamics underlying production decline in high-producing *S. cerevisiae* strains expressing a secreted SARS-CoV-2-mCherry fusion protein. The findings demonstrated that the decline involved the accumulation of mutations within the production plasmid. However, production instability could not be fully explained by genetic changes alone. As shown in **Figure 4.1, Chapter 4**, after repressing the expression for 20 generations, some cultures exhibited partial recovery of the production upon re-induction. This observation indicated that part of the production decline was phenotypically reversible, suggesting that non-genotypic heterogeneity was also involved in shaping the production decline trajectory. Such a pattern supports a model in which early production-associated metabolic burden or stress drives the selection of low-producing phenotypes.

Two key sources of intracellular stress, which are potentially involved in the selection of phenotypic cell variants, are considered to be particularly relevant in the context of recombinant protein production: the unfolded protein response (UPR) and oxidative stress [158; 236]. The UPR is a well-characterised eukaryotic stress pathway that is triggered by the accumulation of misfolded proteins in the endoplasmic reticulum (ER)

[158]. High-level production of secreted or complex proteins, such as the SARS-CoV-2-mCherry fusion protein used in this study, may impose substantial demand on the ER folding machinery. When the capacity to fold and process proteins is exceeded, the UPR is activated to mitigate damage and restore homeostasis through transcriptional upregulation of chaperones, attenuation of general translation and activation of degradation pathways [237; 238]. In yeast, high levels of UPR activation have been shown to impair cellular functions and lead to slow growth [239]. Therefore, high-producing sub-populations that experience elevated UPR stress may grow more slowly than low-producing subpopulations that experience less stress. This differential growth could be one factor driving the selective enrichment of low-producing cells over time, thereby contributing to production instability.

Oxidative stress represents another major intracellular burden in recombinant protein production. Reactive oxygen species (ROS) are generated as byproducts of aerobic metabolism and protein folding processes. Accumulated ROS can damage DNA, protein and lipids, thereby triggering stress responses and impairing cellular performance [240-242]. In microbial cell factories, oxidative stress tolerance is closely linked to production capacity, and enhancing this tolerance has been shown to improve the performance of production strains. For example, yeast strains engineered to overexpress antioxidant-related genes showed improved production of secreted recombinant amylase [243], and strains exhibiting greater resistance to oxidative stress have been identified as better producers of recombinant lysozyme [244].

Besides growth disadvantages caused by metabolic burden and intracellular stress, phenotypic heterogeneity itself impacts production potential. In a heterogeneous

population, only a subset of cells may be in a high-producing state, while others express the recombinant protein at low or undetectable levels. This uneven expression reduces the average yield per cell, meaning that even if some cells achieve high productivity, the population-wide output may be suboptimal. Thus, understanding the extent, dynamics and underlying mechanisms of phenotypic heterogeneity is important for improving both the stability and efficiency of microbial production systems.

5.1.1 Aims

This chapter investigates the role of phenotypic heterogeneity in driving the evolutionary dynamics and production decline observed in the recombinant yeast strains described in Chapter 4. In addition, two key intracellular stresses associated with heterologous protein expression, the UPR and oxidative stresses, are analysed at the single-cell and population level to assess their potential contribution to variation in production. (Single-cell and population-level refer to the resolution of the measurement. Single-cell analyses mean that reporter readouts were recorded from individual cells, allowing the distribution of phenotypes within a culture to be quantified; population-level analyses mean bulk measurements made on cultures that report an average signal across many cells)

Phenotypic heterogeneity within the population was quantitatively measured by the coefficient of variation (CV) between individual cells, as described in the **Section 1.3.2.5**. Fluorescence-activated cell sorting (FACS) was used to isolate distinct subpopulations based on their expression levels from cultures of strains carrying the

initial production plasmids. These subpopulations were subsequently recultured to determine the stability of their production profiles.

5.2 Chapter-specific materials and methods

5.2.1 Strains and cultivation conditions

In this chapter, experiments were conducted using the *S. cerevisiae* retransformants carrying single-type plasmids (High-2_initial, High-2_sp1-6, High-3_initial and High-3_sp1-6) described in Figure 4.4, Chapter 4. These retransformants were *ura3*- and therefore uracil auxotrophs. Unless otherwise stated, all strains and their derivatives were maintained and grown in BMMD + CSM – Leu – Met medium and were generally cultured in 10 mL medium in 50-mL Erlenmeyer flasks at 30 °C, 120 rev. min^{-1} (50 mm orbit).

5.2.2 Fluorescence-activated cell sorting (FACS)

Yeast cells in the exponential growth phase (OD_{620} ~ 1.5 – 2.0) were diluted to an OD_{620} ~ 1.0 using sterile PBS. An aliquot (1 mL) of the diluted sample was washed twice with PBS by centrifugation (1000 x *g*, 2 min) and then resuspended in 1 mL of fresh PBS. Cell sorting was performed using an Astrios Cell Sorter (Beckman Coulter). Based on mCherry fluorescence intensity, three distinct subpopulations were gated: non-producers, low-producers and high-producers. From each gate, 15,000 cells from were collected into sterile PBS for subsequent experiments. In total, ~ 30, 000 cells were analysed per sample and each sorted fraction represented approximately 5% of the overall population .

5.2.3 Construction of unfolded protein response (UPR) reporter system and UPR measurement

A UPR reporter system was constructed using a 98-bp synthetic minimal UPR promoter (*1xUPRE1-SM*) as described by Peng et al. [245]. Primers used were listed in **Table Appendix B.2**. The promoter sequence was synthesised by annealing of the 98-bp forward and reverse oligos: 25 µL of 10 µM forward oligo and 25 µL of 10 µM reverse oligo were mixed, incubated at 98 ℃ for 5 mins and then allowed to cool gradually to room temperature by switching off the heat block. A yeast GFP reporter sequence (*yeGFP-t*ADH1) was amplified from plasmid pSVA12 [246]. The *URA3* selectable marker under control of its native promoter (*pURA3-URA3*) and the *CYC1* terminator (*tCYC1*) were amplified from plasmid pCM190 (EUROSCARF). Additionally, 300-400 bp homology arms corresponding to the upstream (*UP_ura3*) and downstream (*DW_ura3*) of the *URA3* ORF were amplified from *S. cerevisiae* genomic DNA. The *UP_ura3, 1xUPRE1-SM*, *yeGFP-t*ADH1, *pURA3-URA3, tCYC1* and *DW_ura3* fragments were assembled between the *Sac*I and *Sal*I sites of pRS315 via Gibson assembly. The resultant cassette (*UP_ura3-1xUPRE1-SM-yeGFP-t*ADH1-*pURA3-URA3-tCYC1-DW_ura3*) was amplified and used for transformation into yeast strains for UPR measurement. Transformants were selected on YNB agar without the addition of any amino acids. Correct genomic integration was confirmed by PCR and gel electrophoresis using genomic DNA extracted from the transformants as template. UPR activity was measured from yeGFP signal by flow cytometry (see **Section 2.6.1**)

5.2.4 Oxidative stress measurement

Oxidative stress levels in yeast cells were determined using the fluorescent probe 2',7'-dichlorodihydrofluorescein diacetate (H_2DCF-DA), following the protocol described by James *et al*. [247]. Briefly, 1 mL of exponential phase cells (OD_{620} ~ 1.5 – 2.0) was harvested by centrifugation (1000 x *g*, 2 min) and resuspended in 1 mL PBS containing 10 µM H_2DCF-DA. The cell suspension was adjusted to a final OD_{620} ~ 1.0. To prevent premature oxidation of the probe, samples were wrapped in aluminium foil and incubated at 30 °C for 30 minutes with shaking at 120 rev. min^{-1} (50 mm orbit). After incubation, cells were washed twice with 1 mL PBS and resuspended in 1 mL PBS. Fluorescence was measured by flow cytometry as described in the **Section 2.6.1**. The oxidised form of H_2DCF-DA was detected using the B525 channel for green fluorescence (525/40 nm emission filter).

5.3 Results

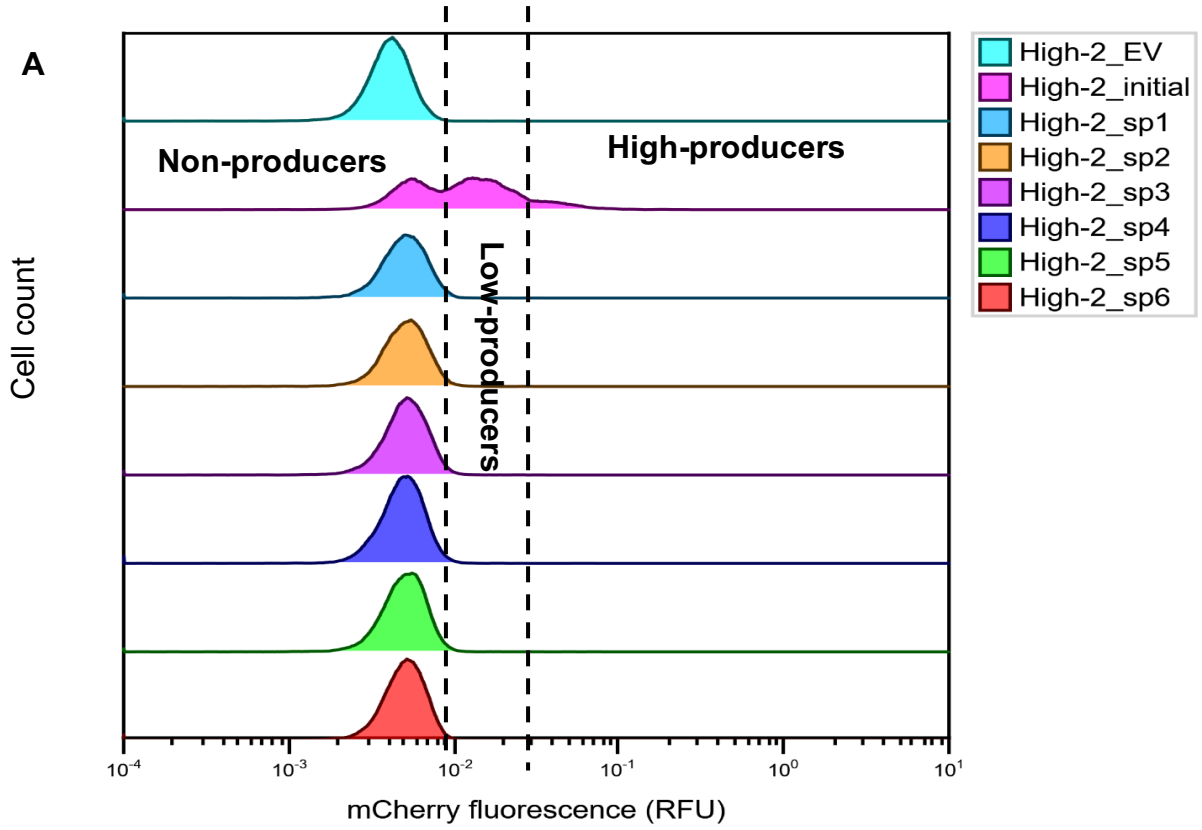
5.3.1 Three distinct subpopulations were observed in strains carrying the initial production plasmids

As shown in **Figure 4.1, Chapter 4**, some cultures exhibited partial recovery of production after growth under non-inducing conditions, suggesting that phenotypic heterogeneity may play a role in loss of production over time, in addition to the plasmid mutations characterised in Chapter 4. Given that individual cells within an isogenic population can differ in gene expression, leading to differences in metabolic activity and stress tolerance [248], it was hypothesised that strains carrying the initial high-production plasmid would exhibit broader distributions of production phenotypes than those later in the production cycle, as low-producing cells may have an advantage and become enriched over time. To investigate this, flow cytometry was used to measure mCherry fluorescence at the single-cell level for High-2 and High-3 strains carrying either the empty vector (High-2_EV and High-3_EV), initial production plasmid extracted from the 10th generation (High-2_initial and High-3_initial) or single-type plasmids isolated from low-producing colonies derived from evolved cultures at 140th generation (High-2_sp1-6 and High-3_sp1-6). The empty vector controls (High-2_EV and High-3_EV) were used to define the fluorescence threshold for non-producers.

As shown in **Figure 5.1**, both High-2 and High-3 strains carrying the initial production plasmid displayed broad, multimodal fluorescence distributions with three apparent subpopulations: a non-producer group with fluorescence level similar to the empty vector control, a group that could be designated ‘low producer’ with intermediate fluorescence intensity, and a group designated ‘high producer’ as it showed elevated mCherry expression. Although the high producer subpopulation was distinguishable,

it comprised only a small fraction of the total cell population (~10%). The majority of cells were either low- or non-producers. This distribution indicates significant phenotypic heterogeneity in the early high-yielding populations.

In contrast, strains carrying the evolved low-producing plasmids derived from the 140th generation displayed markedly narrower fluorescence distributions, predominantly within the non- and/or low-producers ranges. High producing individuals were practically absent, with fewer than 0.1% of cells showing fluorescence intensities comparable to those of the high producer group observed in strains carrying the initial production plasmid. Consistent with this, CV (coefficient of variation) values in mCherry fluorescence were significantly lower for these strains compared to those carrying the initial production plasmid (**Figure 5.2**). Specifically, the High-2 strains carrying the evolved plasmids High-2_sp1-6 showed compact distributions centred around the non-producer range, while the High-3 strains carrying the evolved plasmids High-3_sp1-6, exhibited slightly broader patterns, spanning both non- and low-producer ranges. Interestingly, the High-3 strains carrying High-3_sp1, sp2 and sp4 plasmids, which maintained relatively high expression levels (see **Figure 4.4**), exhibited higher CV values than those carrying High-3_sp3, sp5 and sp6 plasmids (**Figure 5.2**). These results suggest that higher production is associated with greater expression heterogeneity in the cell populations.


A
Cell count
Non-producers
High-producers
Low-producers
High-2_EV
High-2_initial
High-2_sp1
High-2_sp2
High-2_sp3
High-2_sp4
High-2_sp5
High-2_sp6
10-4
10-3
10-2
10-1
100
101
mCherry fluorescence (RFU)

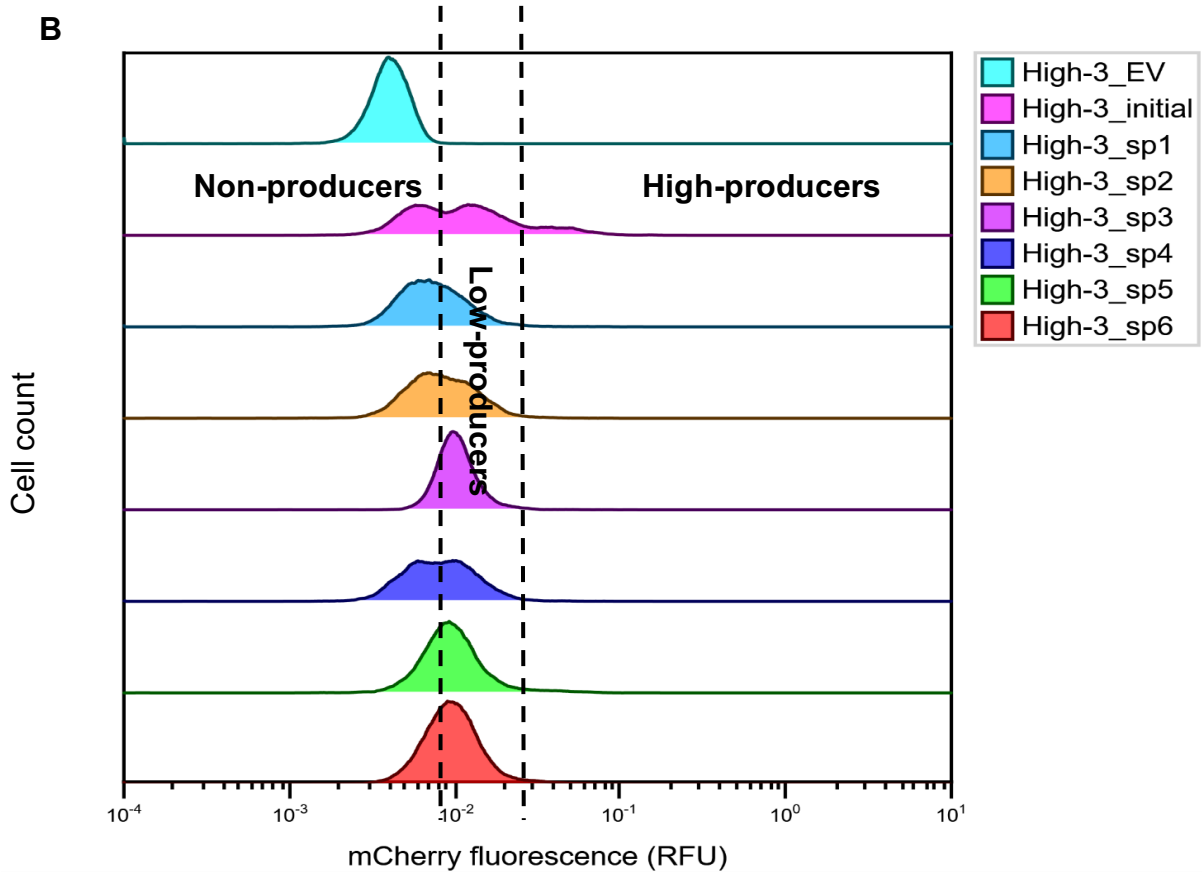

B
Cell count
Non-producers
High-producers
Low-producers
High-3_EV
High-3_initial
High-3_sp1
High-3_sp2
High-3_sp3
High-3_sp4
High-3_sp5
High-3_sp6
10-4
10-3
10-2
10-1
100
101
mCherry fluorescence (RFU)

Figure 5.1: SARS-CoV-2-mCherry fusion protein expression profiles at the single-cell level in High-2 and High-3 retransformants. Flow cytometry analysis of mCherry fluorescence was performed for High-2 (A) and High-3 (B) strains corresponding to those analysed in **Figure 4.4**: High-2 and High-3 strains at the early production stage (10th generation) cured of plasmids were retransformed with either an empty vector (High-2_EV and High-3_EV), initial production plasmids extracted from the 10th generation cultures (High-2_initial and High-3_initial) or plasmids extracted from low-producing colonies derived from evolved cultures at 140th generation (High-2_sp1-6 and High-3_sp1-6). Dashed vertical lines indicate approximate fluorescence thresholds distinguishing 'non-producers', 'low-producers' and 'high-producers', defined based on the fluorescence distribution of the empty vector control (non-producers) and the tripartite profile observed in strains carrying the initial production plasmids. Each fluorescence distribution represents data from 10,000 individual cells, with mCherry fluorescence values normalised to cell size (FSC-A). mCherry fluorescence is reported as RFU (relative fluorescence units, fluorescence normalised as indicated).

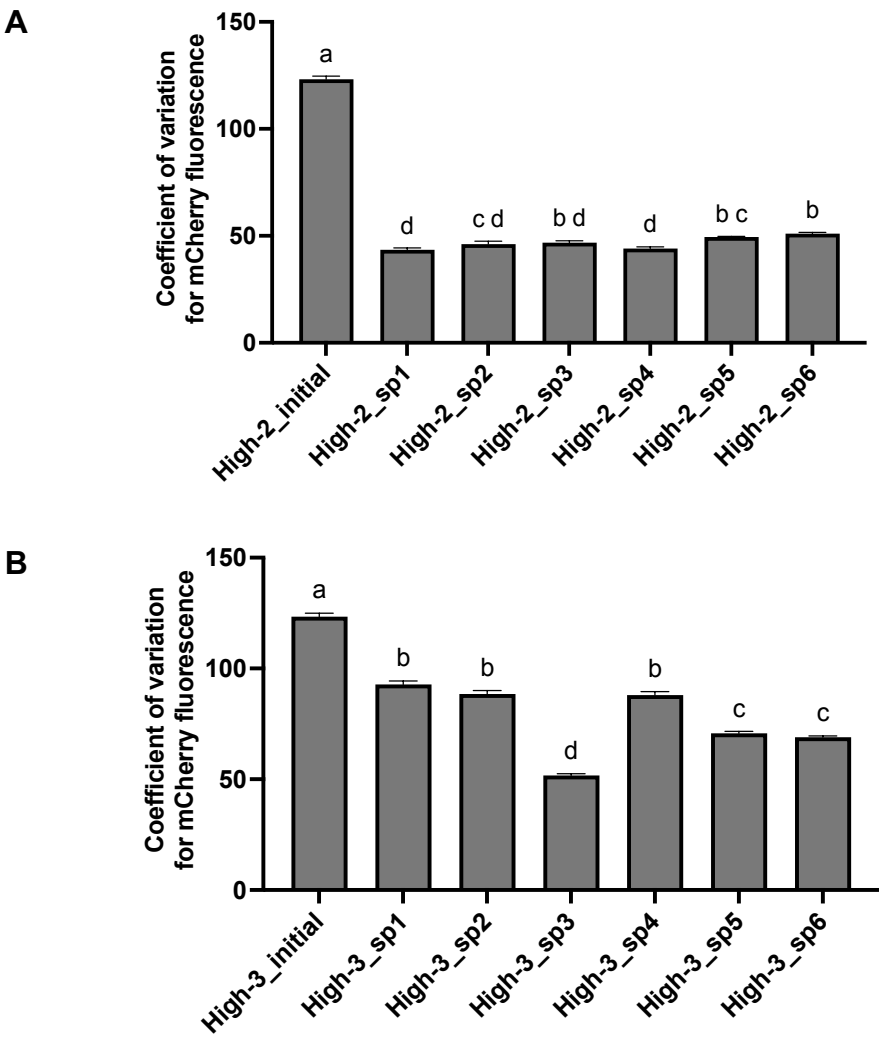


Figure 5.2: Heterogeneity level of mCherry fluorescence in High-2 and High-3 strains carrying the initial production plasmid or the evolved low-producing plasmids, measured by coefficient of variation. Cell-to-cell heterogeneity level in SARS-CoV-2-mCherry fusion protein expression was assessed using the coefficient of variation (CV) of single-cell mCherry fluorescence for High-2 (A) and High-3 (B) strains carrying either the initial production plasmid (High-2_initial and High-3_initial) or the evolved low-producing plasmids (High-2_sp1-6 and High-3_sp1-6). Data were obtained via flow cytometry. Bars represent means ± SEM of three biological replicates, each biological replicate based on measurements from 10,000 cells. Means with different letters are significantly different (ordinary one-way ANOVA: (A) $F(6, 14) = 906.5$, $p < 0.0001$; (B) $F(6, 14) = 314.0$, $p < 0.0001$).

5.3.2 The production phenotypes of the three subpopulations were transient, indicating non-heritable

To determine whether the three subpopulations observed in strains carrying the initial production plasmids were driven by phenotypic heterogeneity rather than any arising genetic differences, fluorescence-activated cell sorting (FACS) was used to isolate approximately 15,000 cells from each of the non-producing, low-producing and high-producing subpopulations. These sorted subpopulations were then independently regrown in BMMD + CSM – Leu – Met medium for three days and their expression levels measured by plate reader. Strikingly, the expression levels of the three subpopulations showed no significant differences and resembled the average expression level of the original unsorted population prior to sorting (**Figure 5.3**). This suggests that the production phenotypes of the three subpopulations were transient and not heritable. Flow cytometry analysis also revealed that each regrown subpopulation re-established a broad, multimodal fluorescence distribution similar to that of the original culture, encompassing all three expression states (**Figure 5.4A**). These findings indicate that the observed subpopulations represent reversible phenotypic variation rather than stable heritable traits.

One potential concern is that the observed re-distribution could have been an artefact of poor sorting resolution, with insufficient or unsuccessful FACS discrimination between the gated subpopulations. However, post-sort reanalysis was conducted immediately after sorting and confirmed that the sorted populations were highly enriched for cells within their respective fluorescence ranges defined by the gating

(**Figure 5.4B**). This supports the reliability of the sorting procedure and strengthens the conclusion that broadened phenotype distribution occurred during subsequent outgrowth.

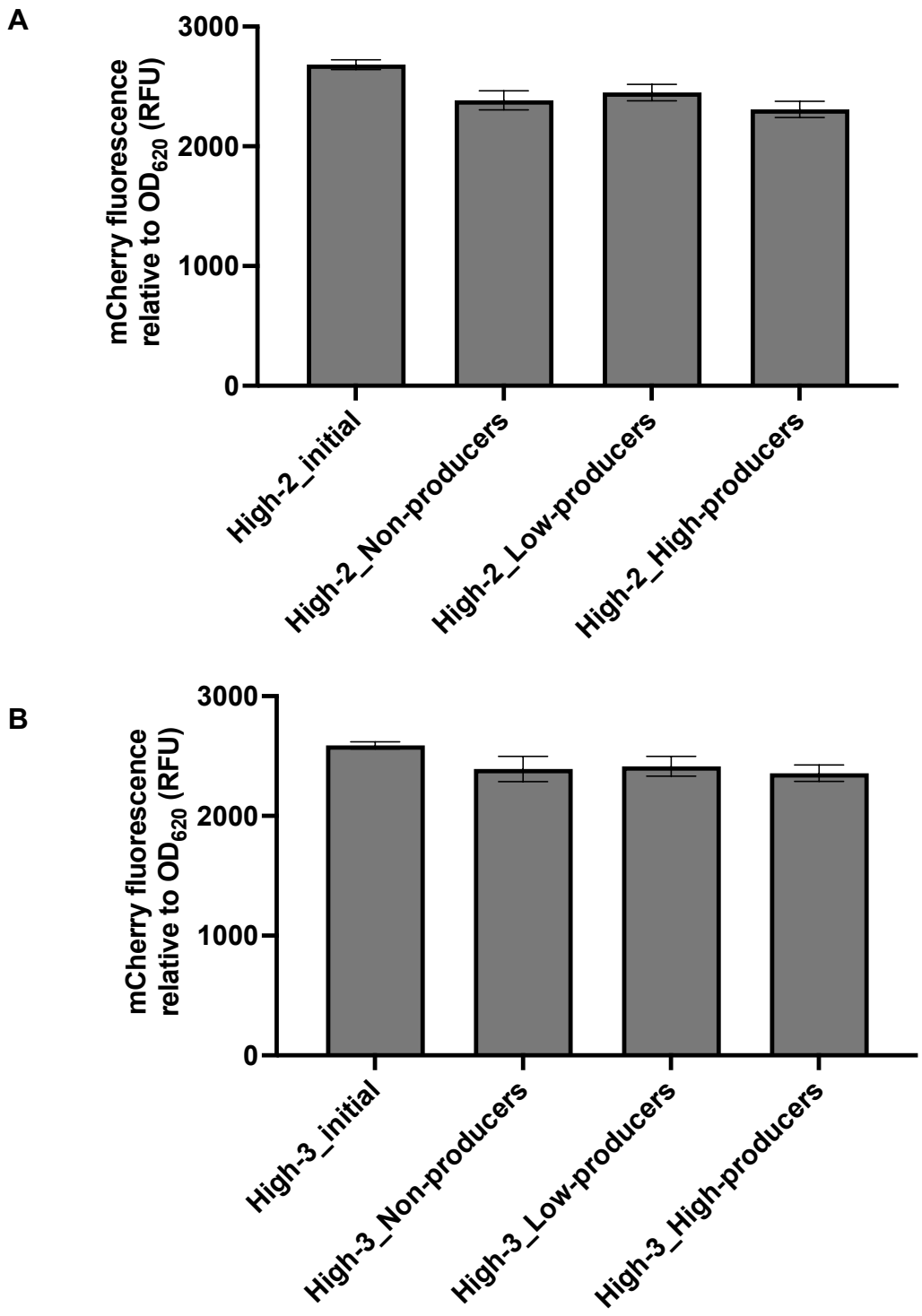


Figure 5.3: SARS-CoV-2-mCherry expression levels following regrowth of sorted subpopulations. Expression levels of the SARS-CoV-2-mCherry fusion protein were measured in sorted subpopulations of High-2 (A) and High-3 (B) carrying the initial production plasmid. The original unsorted cultures (High-2_initial and High-

3_initial) were used as controls. Approximately 15,000 cells were sorted from each of the non-producing, low-producing and high-producing subpopulations based on mCherry fluorescence and cultured in BMMD + CSM – Leu – Met medium for three days. Fluorescence intensity was then quantified using a plate reader and normalised to OD_{620}. Data represent mean values ± SEM of six biological replicates. Statistical significance was assessed by ordinary one-way ANOVA: (A) $F(3, 20) = 2.783$, $p = 0.0675$ (ns). (B) $F(3, 20) = 1.821$, $p = 0.1759$ (ns).

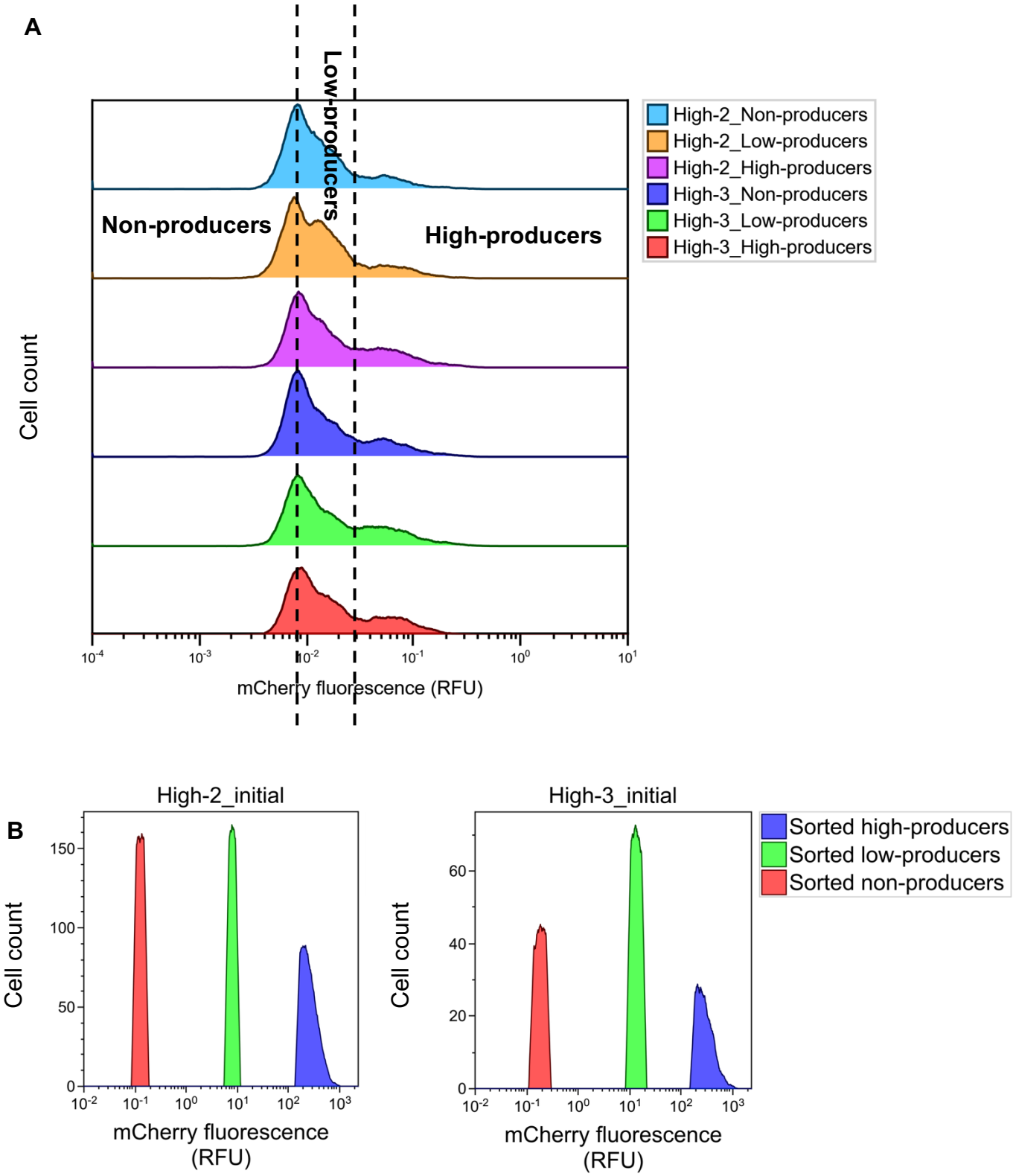


Figure 5.4: SARS-CoV-2-mCherry fusion protein expression at single-cell level following cell sorting and regrowth of subpopulations. Approximately 15,000 cells were isolated from each subpopulation (non-producers, low-producers, and high-

producers) of High-2 and High-3 strains carrying the initial production plasmid based on mCherry fluorescence normalised to cell size (FSC-A). mCherry fluorescence is reported as RFU (relative fluorescence units, fluorescence normalised as indicated). (A) Sorted cells were regrown in BMMD + CSM – Leu – Met medium and their SARS-CoV-2-mCherry fusion protein expressions were re-analysed by flow cytometry (CytoFLEX S Flow Cytometer). Each distribution represents data from 10,000 individual cells. (B) Post-sort reanalysis performed immediately after sorting (Astrios Cell Sorter) confirmed that the sorted populations were highly enriched within their respective fluorescence ranges defined by gating. Note: data in panels A and B were obtained using different flow cytometers, and mCherry values are therefore not directly comparable.

5.3.3 Correlation between UPR activation and production at the single-cell level in strains carrying the initial, high-production plasmid

Previous studies have shown that phenotypic heterogeneity is prevalent in cellular stress resistance or stress responses, where it may function as an adaptive strategy that enables microbial populations to persist through potentially transient environmental or intracellular challenges [214; 249-251]. In the context of recombinant protein production, there may be heterogeneity in the level of stress experienced by individual cells. Specifically, cells with lower production burdens, such as the non- and low-producing subpopulations examined above, may experience reduced intracellular stress and therefore exhibit greater fitness under production conditions. By contrast, high-producing subpopulations likely suffer elevated stress, which may compromise their growth and competitiveness. Over time, those high-producing subpopulations

may be selectively outcompeted and gradually lost from the population, consistent with some of the above results.

Recombinant protein secretion imposes a high metabolic burden, particularly through energy-intensive processes such as folding, glycosylation and translocation [252]. To examine whether intracellular stress responses might contribute to the emergence of the three distinct subpopulations described in **Section 5.3.1** and potentially be a driver of the changes in production observed during long-term cultivation. the UPR was analysed. The UPR is the primary stress response pathway in yeast associated with protein folding and secretion stress and is therefore directly relevant to secretion of the SARS-CoV-2-mCherry fusion protein [253].

To quantify UPR stress, a reporter system based on a 98-bp synthetic minimal UPR sensor developed by Peng et al. was employed [245]. Briefly, this synthetic promoter element, referred to as *1xUPRE1-SM*, was designed from conserved upstream regulatory sequences of canonical UPR target genes and is specifically responsive to UPR activation. In this study, the *1xUPRE1-SM* element was cloned upstream of a yeGFP reporter, and the resulting reporter cassette (*1xUPRE1-SM-yeGFP-t*ADH1) was integrated at the *URA3* locus in the High-2 and High-3 strains carrying either the initial production plasmid or the evolved low-producing plasmids (see **Section 5.2.3**). This enabled quantification of UPR activity based on yeGFP fluorescence, allowing direct comparison of UPR activation levels between different production states at both the single-cell and population levels.

As expected, single-cell data revealed a general trend in which higher SARS-CoV-2-mCherry fusion protein production was associated with greater UPR activation in both High-2 and High-3 strains carrying the initial production plasmid (**Figure 5.5**). This observation supports the hypothesis that cells producing at high levels experience greater UPR stress. The relationship between production and UPR activation was not strictly linear. The coefficient of determination (R^2) for the relationship between mCherry fluorescence (as a proxy for production) and yeGFP fluorescence (as a proxy for UPR activation) was 0.4608 for High-2 and 0.3585 for High-3 strains. This suggests that approximately 46% and 36% of the variation in UPR activation, respectively, could be explained by differences in production levels. In both cases, a secondary, shallower trendline of datapoints was also visible on the plots (**Figure 5.5**), which may reflect additional subpopulation behaviour or technical factors during flow cytometry measurement.

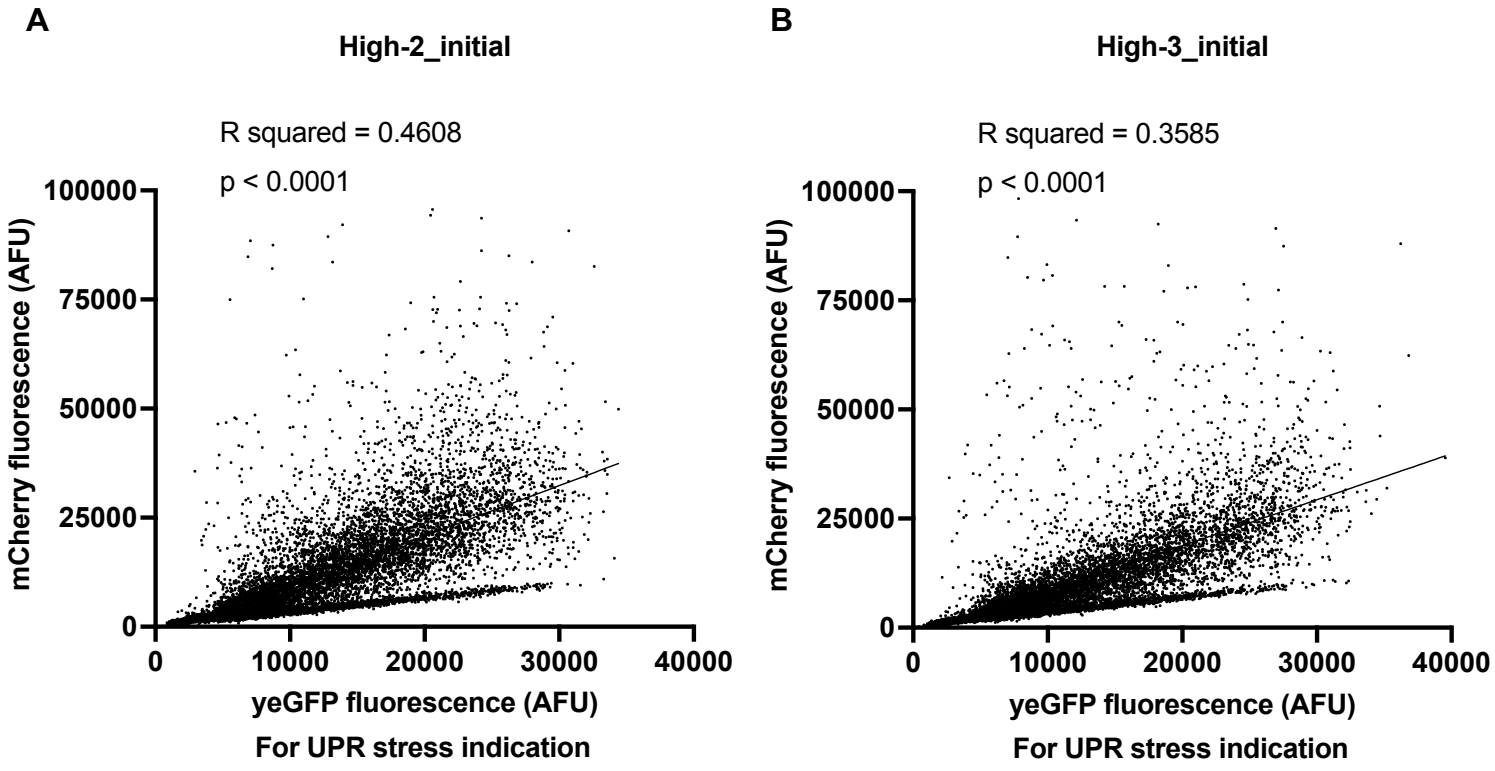


Figure 5.5: Correlation between UPR activation and production in High-2 and High-3 strains carrying the initial production plasmid at the single-cell level. UPR activation levels were quantified using a yeGFP-based reporter system driven by a synthetic minimal UPR-inducible promoter. mCherry and yeGFP fluorescences are reported as AFU (arbitrary fluorescence units, instrument output). Panels A and B show single-cell UPR reporter signals (yeGFP fluorescence) plotted against SARS-CoV-2-mCherry fusion protein expression levels for High-2 (A) and High-3 (B) strains carrying the initial production plasmid. Each dot represents an individual cell (n = 10,000), measured by flow cytometry. The solid line represents a simple linear regression fit.

5.3.4 Similar population-level UPR activation across strains, but elevated heterogeneity in high producers.

Next, to assess whether UPR activation differed between strains carrying the initial high-production plasmid (High-2_initial and High-3_initial) and those carrying evolved low-production plasmids (High-2_sp1-6 and High-3_sp1-6), population-level UPR activation was measured using the yeGFP-based UPR reporter via flow cytometry. Contrary to expectation, the mean UPR activation, quantified as yeGFP fluorescence normalized to cell size, did not differ significantly between strains carrying the initial production plasmid and those carrying evolved plasmids (**Figure 5.6A and B**). This contrasted with the prior single-cell analysis (**Figure 5.5**), where higher SARS-CoV-2-mCherry production correlated with stronger UPR activation of individual cells.

In contrast, substantial differences were observed in the heterogeneity of UPR activation, as quantified by the CV in yeGFP fluorescence (**Figure 5.6C and D**). Strains carrying the initial high-production plasmid showed markedly greater variability in UPR activation than those carrying evolved low-producing plasmids. This observation is consistent with the broader production heterogeneity observed previously (**Figure 5.1**). This trend was particularly evident in the High-3 background (**Figure 5.6D**): the High-3_initial strain exhibited the highest UPR heterogeneity, while among the evolved plasmid strains, those with relatively higher production levels (High-3_sp1, sp2 and sp4) also showed greater UPR variability than strains carrying plasmids associated with lower production (High-3_sp3, sp5 and sp6). These findings suggest that while overall UPR activation at the population level may not always reflect differences in production state, heterogeneity in stress response does.

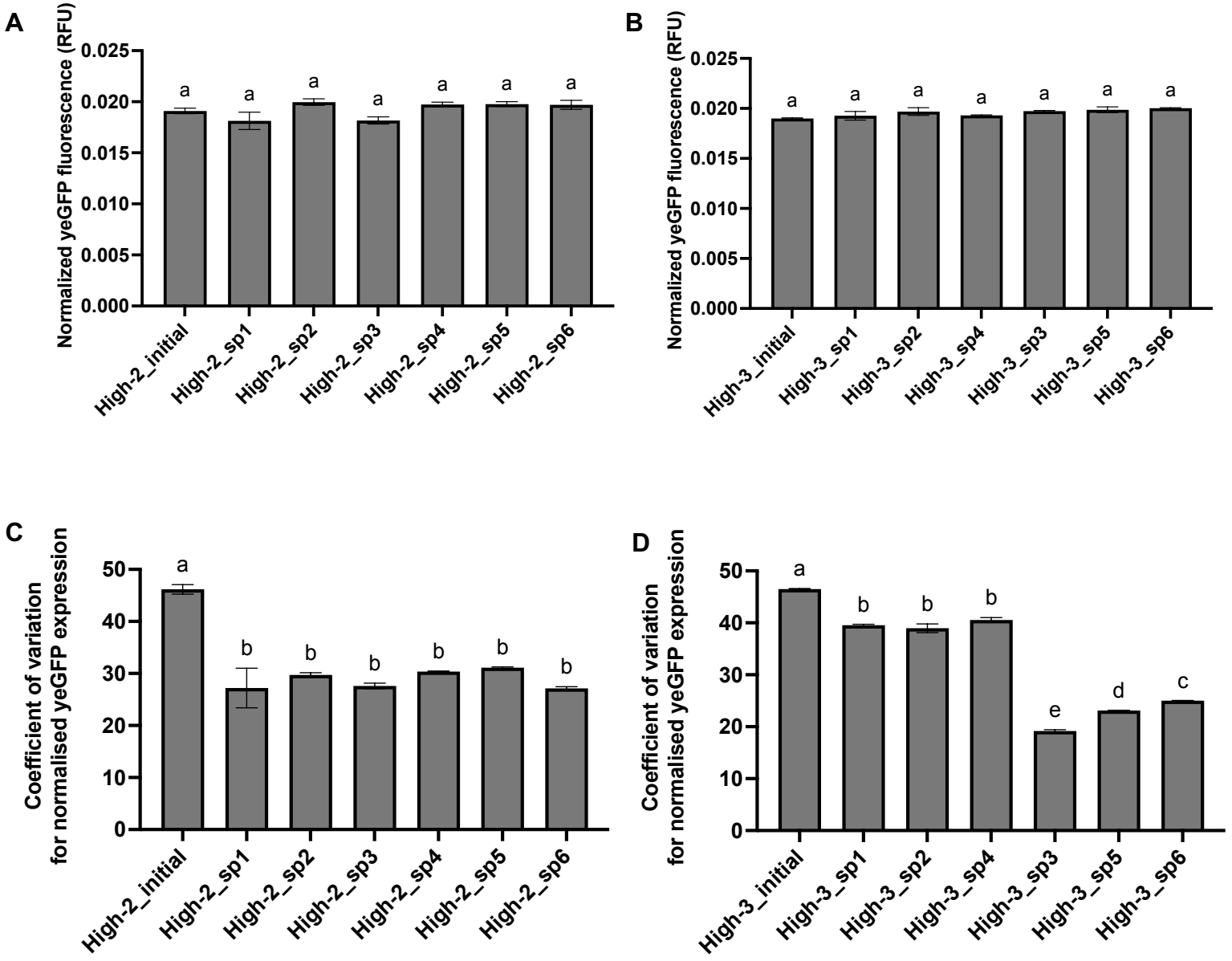


Figure 5.6: Population-level UPR activation and heterogeneity in High-2 and High-3 strains carrying initial or evolved production plasmids. Panel A and B show population-level UPR activation, quantified by yeGFP fluorescence, in High-2 (A) and High-3 (B) strains carrying either the initial production plasmid or evolved low-producing plasmids. Panel C and D present the corresponding heterogeneity in UPR activation, measured as CV of yeGFP fluorescence, in High-2 (C) and High-3 (D) strains. Fluorescence signals were collected from 10,000 cells per sample via flow cytometry and normalised to cell size (FSC-A) and reported as RFU (relative

fluorescence units, fluorescence normalised as indicated). Bars represent means ± SEM of three biological replicates. Means with different letters are significantly different (Ordinary one-way ANOVA: (A) $F(6, 14) = 3.250$, $p = 0.0324$; (B) $F(6, 14) = 2.152$, $p = 0.1114$ (ns); (C) $F(6, 14) = 20.04$, $p < 0.0001$; (D) $F(6, 14) = 770.9$, $p < 0.0001$).

5.3.5 Correlation between oxidative stress and production at the single-cell level in strains carrying the initial high-production plasmid

In addition to the UPR, oxidative stress is another key intracellular stress that may impact high-producing populations. Protein folding and secretion are known to generate ROS as by-products and consume reducing agents such as glutathione, thereby increasing the oxidative burden in cells [254]. As a result, high-producing cells may be more susceptible to oxidative stress due to their elevated demand on the ER folding machinery and redox homeostasis. Building on the previous findings that linked high production with increased UPR activation and heterogeneity, it was hypothesised that a similar trend may be observed for oxidative stress.

To assess intracellular oxidative stress, cells were incubated with H_2DCF-DA, and green fluorescence intensity was quantified by flow cytometry (see **Section 5.2.4** for details). As expected, single-cell data from High-2 and High-3 strains carrying the initial production plasmid revealed a clear positive correlation between oxidative stress and SARS-CoV-2-mCherry expression (**Figure 5.7**). This trend mirrors that observed for UPR activation (**Figure 5.5**) and supports the hypothesis that high-producing cells suffer a greater oxidative burden. However, in contrast to the UPR data, where a

secondary, shallower trendline was apparent (**Figure 5.5**), the oxidative stress plots showed a single dominant linear relationship. This absence of a visible secondary subpopulation may explain why the correlation appeared stronger for oxidative stress, with R^2values of 0.5871 for High-2 and 0.6416 for High-3 strains. These values suggest that approximately 59% and 64% of the variation in oxidative stress in High-2 and High-3 strains, respectively, could be explained by differences in production levels.

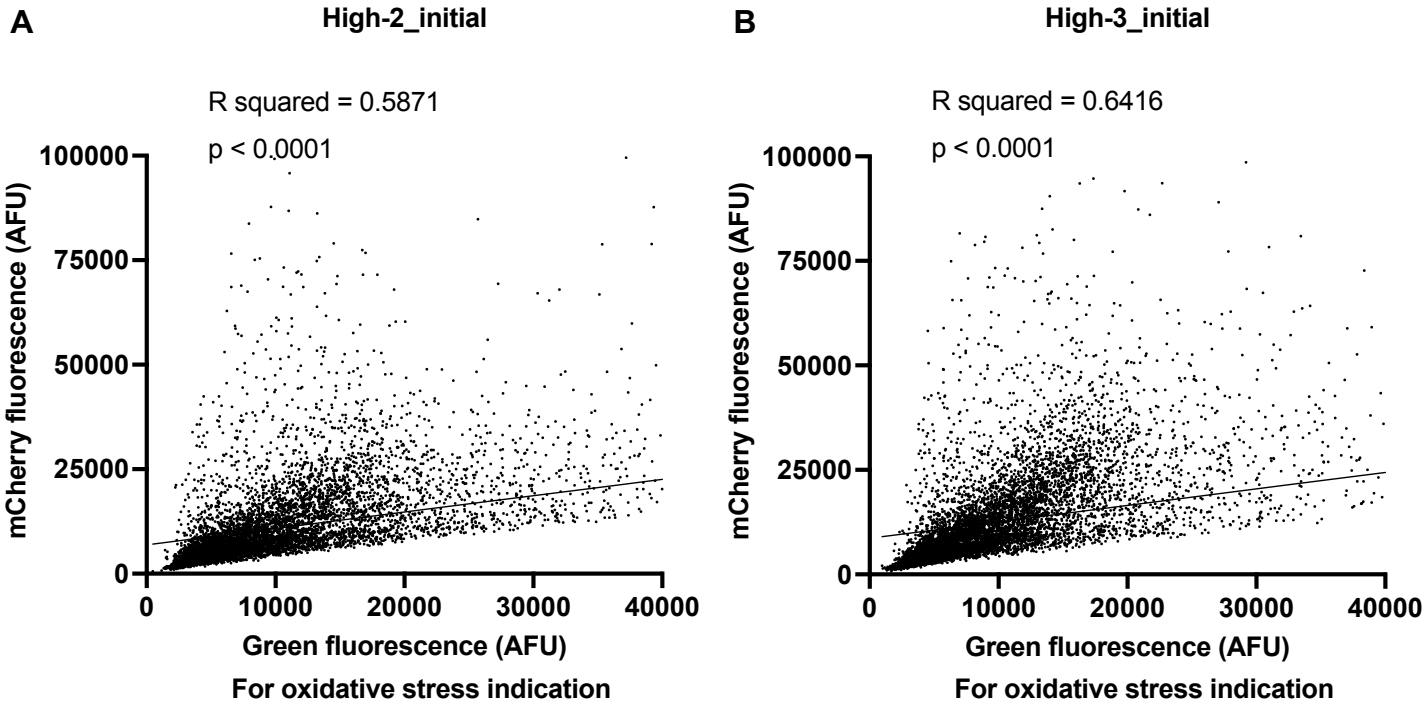


Figure 5.7: Correlation between oxidative stress and production at the single-cell level in High-2 and High-3 strains carrying the initial production plasmid.

Oxidative stress levels were measured using the fluorescent dye H_2DCF-DA. Exponential phase cells were harvested and resuspended in 1 mL PBS containing 10 µM H_2DCF-DA and incubated at 30 °C for 30 min. mCherry and yeGFP fluorescences are reported as AFU (arbitrary fluorescence units, instrument output). Panels A and B

show single-cell oxidative stress levels (green fluorescence) plotted against SARS-CoV-2-mCherry fusion protein expression levels (mCherry fluorescence) in High-2 (A) and High-3 (B) strains carrying the initial production plasmid. Each dot represents an individual cell (n = 10,000), measured by flow cytometry. The solid line represents a simple linear regression fit.

5.3.6 Elevated population-level oxidative stress and heterogeneity in high producers

To investigate whether oxidative stress differed between strains carrying the initial high-production plasmid (High-2_initial and High-3_initial) and those carrying evolved low-production plasmids (High-2_sp1-6 and High-3_sp1-6), population-level oxidative stress was measured. In contrast to UPR activation (**Figure 5.6A and B**), oxidative stress measurements revealed significant differences in mean fluorescence intensity between the two groups (**Figure 5.8A and B**). Strains carrying the initial high-production plasmids exhibited higher oxidative stress levels, consistent with the elevated production burden and strong positive correlation observed at the single-cell level (**Figure 5.7**). Moreover, the high-producing strains also exhibited substantially greater heterogeneity in oxidative stress, as indicated by higher CV values in green fluorescence (**Figure 5.8C and D**). The strains carrying the initial production plasmid (High-2_initial and High-3_initial) displayed the highest oxidative stress variability. Among the strains carrying the evolved plasmids, those with relatively higher production (High-3_sp1, sp2 and sp4) also showed intermediate levels of oxidative stress heterogeneity. The results compared to those for UPR suggest that oxidative stress could be a more sensitive and reliable indicator of recombinant protein

production burden, at least in the context of SARS-CoV-2-mCherry fusion protein expression.

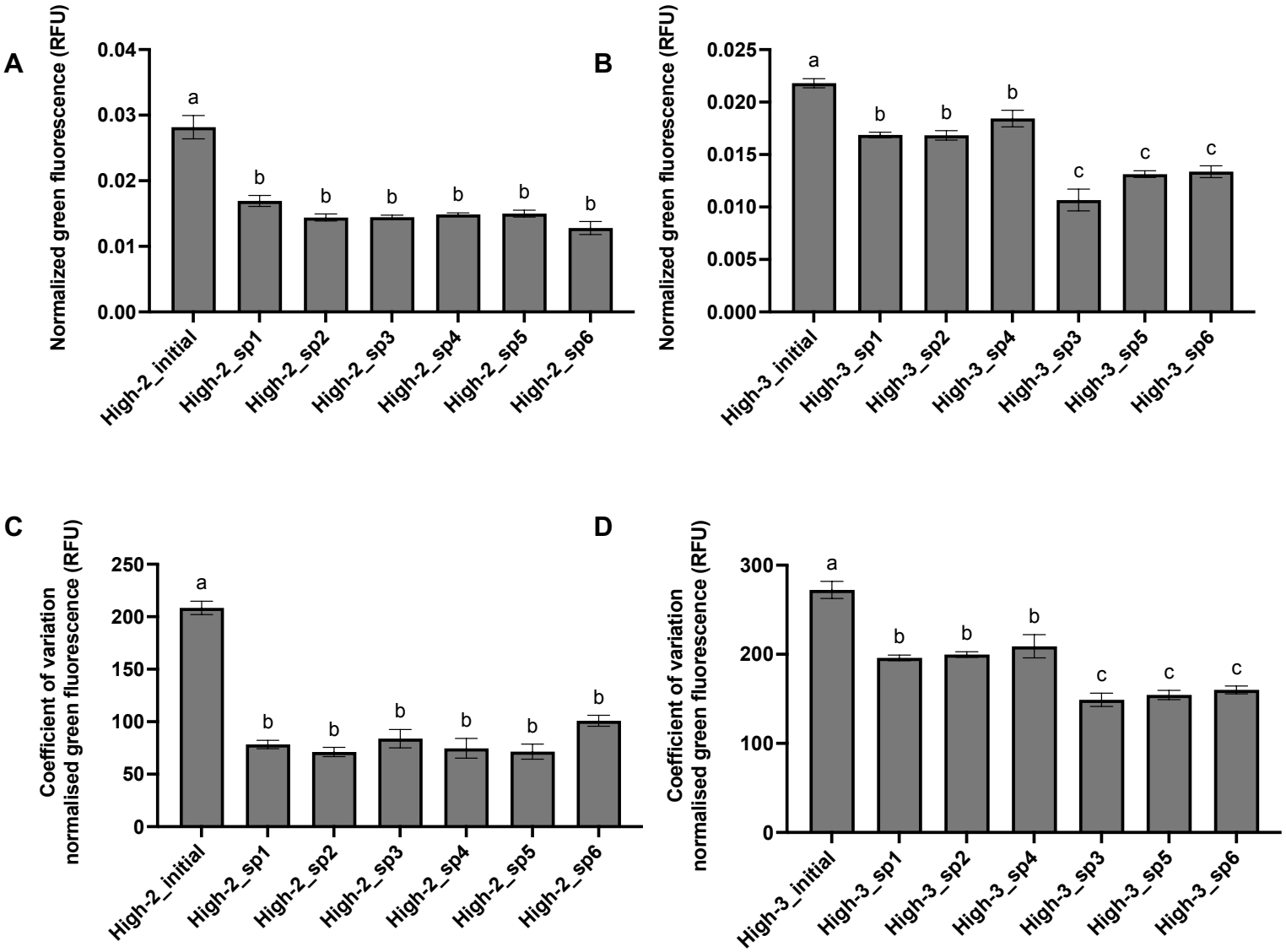


Figure 5.8: Population-level oxidative stress and heterogeneity in High-2 and High-3 strains carrying initial or evolved production plasmids. Panel A and B show population-level oxidative stress, quantified by green fluorescence, in High-2 (A) and High-3 (B) strains carrying either the initial production plasmid or evolved low-producing plasmids. Panel C and D present the corresponding heterogeneity in oxidative stress, measured as CV of green fluorescence, in High-2 (C) and High-3 (D)

strains. Fluorescence signals were collected from 10,000 cells per sample via flow cytometry and normalised to cell size (FSC-A) and reported as RFU (relative fluorescence units, fluorescence normalised as indicated). Bars represent means ± SEM of three biological replicates. Means with different letters are significantly different (Ordinary one-way ANOVA: (A) $F(6, 14) = 34.30$, $p < 0.0001$; (B) $F(6, 14) = 38.06$, $p < 0.0001$; (C) $F(6, 14) = 53.80$, $p < 0.0001$; (D) $F(6, 14) = 33.78$, $p < 0.0001$).

5.4 Discussion

This chapter investigated the contribution of phenotypic heterogeneity to production instability in high-producing yeast strains. Sing-cell analyses revealed the coexistence of non-, low- and high-producing subpopulations within isogenic cultures, and cell sorting following regrowth of those subpopulations demonstrated that these states were phenotypically reversible rather than genetically fixed. High-producing cells were consistently associated with elevated intracellular stress, with oxidative stress showing a stronger correlation with production levels than UPR activation in the case of SARS-CoV-2-mCherry fusion protein expression. At the population level, strains carrying the initial high-production plasmids displayed both higher average oxidative stress and greater heterogeneity compared with those carrying the evolved, low-production plasmids.

5.4.1 Phenotypic heterogeneity and its link to production instability

At the initial stage of the protein production experiment described here, three distinct subpopulations, non, low and high-producers were identified before the emergence of genetic mutations. This observation suggested that production-associated stress could contribute to heterogeneous expression at the single-cell level, although stochastic variation and other sources of phenotypic noise are also likely to play a role. Similar phenotypic distributions, in which yeast populations split into bimodal or trimodal expression or metabolic activity states, have previously been reported in yeast under nutrition-limited conditions, and such heterogeneity was interpreted as an adaptive strategy to enhance population survival [77; 133; 139; 255-258]. These studies, together with the findings in this chapter that strains carrying the low-producing plasmids exhibited diminished heterogeneity, suggest that phenotypic

heterogeneity is a common feature of yeast populations and can be influenced by biosynthetic burden and stress. Such heterogeneity may provide flexibility that enables populations to cope with environmental or metabolic fluctuations, although its extent and persistence vary depending on the metabolic and environmental conditions.

Importantly, when the three distinct subpopulations were sorted and regrown, the original trimodal distribution reappeared. This recurrence strongly suggests that these expression states were not fixed by genetic factors but instead represented reversible phenotypic states. Some degree of contamination during cell sorting cannot be entirely excluded, for example, a very small fraction of cells may be mis-sorted due to cell clumping, but this is unlikely to account for the consistent re-emergence of all three states. Future validation could be obtained using complementary approaches, such as lineage-tracking or microfluidics-based sing-cell cultivation, which allow more precise monitoring of state transitions over time [259].

The High-3 strain carrying the initial, unmutated production plasmid exhibited greater cell-to-cell variability in the protein expression compared to the High-3 strains carrying the evolved, mutated plasmids. Among the strains carrying the evolved plasmids, those harbouring relatively higher-producing evolved plasmids (i.e. High-3_sp1, sp2 and sp4) displayed more heterogeneous expression profiles than those carrying the low-producing plasmids (i.e. High-3_sp3, sp5 and sp6). This pattern suggests that elevated phenotypic heterogeneity was associated with higher production levels and might represent a transient response to the production burden. If the High-3 strain was cultivated for more than 140 generations, it seems reasonable to anticipate that both production levels and expression heterogeneity may further decline, eventually

reaching a more uniform, low-expression state similar to that observed in the High-2 strain.

Moreover, only around 10% of the population exhibited high-level production with the initial production plasmid, while the majority were non- or low-producers, reflecting substantial cell-to-cell variation in expression. This distribution highlights a key challenge in microbial biomanufacturing: even in genetically optimised strains, non-genetic factors such as phenotypic heterogeneity might limit the proportion of productive cells, capping overall yields. Addressing this bottleneck will require the development of strategies to stabilise or support the high-producing phenotype at the single-cell level. These may include designing systems that couple single-cell production with single-cell fitness in a given condition.

5.4.2 Intracellular stress as a driver of phenotypic heterogeneity

In addition to identifying distinct production subpopulations, this chapter also explores the role of intracellular stress in shaping phenotypic heterogeneity during recombinant protein production. Two key stress responses, UPR and oxidative stress, were analysed in relation to protein production levels. Both stress markers showed positive correlation with production, but oxidative stress demonstrated a stronger association. This difference in correlation strength suggests that oxidative stress may be a more direct or sensitive indicator of production burden than UPR activation, at least within the experimental context of this thesis involving SARS-CoV-2-mCherry fusion protein expression. One possible explanation is that ROS generation is a more immediate by-product of increased protein folding and secretion activity, whereas UPR activation involves a more complex regulatory cascade that may be influenced by additional

cellular factors. Importantly, neither oxidative stress nor UPR levels exhibited a strictly linear relationship with production, implying that additional factors other than production levels contributed to the observed variability in the stress responses. For example, differences in ER folding capacity between cells and heterogeneity in single-cell stress resistance could modulate stress responses independently of production level [260]. Previous work has shown that ROS can inhibit UPR signalling in yeast, which may explain why the correlation between oxidative stress and production level was stronger [261]. This interplay also suggests that stress pathways do not operate independently, but instead interact in complex, sometimes antagonistic ways, further shaping the production phenotype at the single-cell level.

It was hypothesised that, at the population level, strains carrying evolved low-producing plasmids would exhibit both reduced stress levels and lower heterogeneity in stress levels. This pattern was indeed observed for oxidative stress. However, UPR activation did not follow the same trend. While strains carrying evolved low-producing plasmids generally showed lower oxidative stress, their UPR levels were comparable to those observed in strains harbouring the initial high-producing plasmid. Nevertheless, the heterogeneity of UPR activation aligned with expectations: high-producing strains displayed greater variability in UPR signalling, supporting that elevated production levels are linked to increased heterogeneity in intracellular stress responses. A possible explanation is that in high-producing populations, only a small fraction of cells (i.e. those high-producing subpopulations) bears the full production burden and thus activate the UPR strongly, leading to greater population-level variability but not necessarily a higher mean signal. In contrast, low-producing strains likely consist of more uniformly low-stressed cells, leading to lower heterogeneity in

stress. Additionally, evolved low-producing plasmids were shown to have higher copy numbers than the initial production plasmid (see **Figure 4.8**), which may influence the expression of other plasmid-encoded genes such as the selection marker or FLP recombinase, both of which could affect cellular stress levels and UPR activation.

The pronounced heterogeneity in oxidative stress observed among high-producing populations likely influences population dynamics during long-term cultivation. High-producing cells experiencing higher oxidative stress may suffer reduced fitness, while non- or low-producing cells, which experience less oxidative burden, maintain a competitive advantage [170]. Over time, this imbalance may drive a selective shift in the population towards less stressed, non- and/or low-producing phenotypes.

5.4.3 Conclusions and future work

The results of this chapter indicate that phenotypic heterogeneity is another important factor contributing to production instability in high-producing yeast strains. Three distinct subpopulations, non-, low- and high-producers, were identified in the initial production cultures, and their production states were found to be transient and reversible, indicating a non-genotypic origin. This heterogeneity effectively reduces the proportion of productive cells within the population at the beginning, thereby limiting the overall production yield. High-producing cells exhibited elevated UPR and oxidative stress levels, with oxidative stress showing the stronger correlation with production. These findings suggest that oxidative stress could play an important role in the selective loss of high-producing cell subpopulations during long-term cultivation.

However, it is important to recognise that recombinant protein production imposes multiple intracellular stresses. While UPR showed weaker correlation with production levels, unfolded protein stress still likely contributed to cellular burden. Other stress pathways may also be involved. For example, secretory pathway dysfunction caused by imbalances between protein folding and disulfide bond formation has been reported to trigger osmotic stress responses [254]. Future studies could employ transcriptomic or proteomic profiling to help dissect these stress responses and their potential interplay with production heterogeneity, or test additional candidate stresses in a similar way to the approaches used here for UPR and oxidative stress.

To further investigate the mechanisms underlying stress-induced heterogeneity and inform strategies for improving production stability, future work could include single-cell transcriptomics on sorted subpopulations to identify stress and metabolic pathways that differentiate non-, low- and high-producing cells [262], thereby potentially uncovering key regulators of stress tolerance and productivity. Microfluidic-based time-lapse imaging may provide insights into the dynamics and reversibility of phenotypic switching at the single-cell level [259]. In addition, enhancing antioxidant capacity, such as through overexpression of glutathione biosynthesis genes or supplementation of glutathione (or other antioxidants) directly in the growth medium, could help buffer oxidative stress in high-producing cells [263] and potentially improve long-term production stability. Early-stage screening of strain libraries for low expression heterogeneity may provide a valuable strategy for identifying robust production hosts. Besides, active approaches to reduce heterogeneity may also be valuable, including engineering regulatory circuits to promote more uniform expression or applying continual selection strategies that enrich high producers [145; 213]. Overall,

understanding and mitigating phenotypic heterogeneity will be important for achieving stable, high-level expression of recombinant proteins in industrial bioprocesses.

6 EXPLOITING PHENOTYPIC HETEROGENEITY TO IMPROVE PRODUCT YIELD

6.1 Introduction

As described in Chapter 5, phenotypic heterogeneity is common in high-producing yeast populations and can contribute to production instability. At the beginning of cultivation, individual cells within a genetically uniform cell population may display a broad range of product titres, leading to lower overall yield than if all cells were producing maximally. This variability between cells has been reported not only in recombinant protein production but also in other microbial bioprocesses, such as the production of free fatty acid and tyrosine [145; 264-266]. In such cases, only a subset of cells (the high-producing subpopulation) operates near the optimal production state.

Although phenotypic heterogeneity can be considered as a problem in bioproduction, it may also offer an opportunity: if high-producing cells can be selectively enriched, it may be possible to improve yield at the population level. In *E. coli,* for example, improved production of free fatty acid and tyrosine was achieved by using product-activated promoters to confer a growth advantage to phenotypically high-performing cells under antibiotic selection [145]. However, very few studies have attempted to exploit phenotypic heterogeneity for enhanced biosynthesis in yeasts. In two recent examples, biosensor-based selection systems were used to extend productive lifetimes of heterologous vanillin-*β*-glucoside and *N*-acetylglucosamine in *S. cerevisiae* via selection of producing subpopulations [79; 267]. In one case, an essential gene was placed under the control of a synthetic promoter inducible in the presence of the vanillin-*β*-glucoside precursor [267]. In the other, a conditional

addiction circuit was engineered using the suicide gene *FCY1*, whose transcripts were degraded by a *glmS* ribozyme in response to a high level of the product intermediate [79]. Although both systems helped stabilise production over a longer period, the production rate under selection did not surpass the original non-selected control rates. This might possibly be because the engineered control circuits responded to the target product intermediate rather than the final product, limiting any improvement in production rate beyond initial control levels. By contrast, the systems used in *E. coli* responded directly to final product levels and achieved up to a three-fold increase in the production of both free fatty acid and tyrosine, compared to the original level of the non-selected controls [145]. Applying this type of selection in yeast, where the regulatory response is linked directly to the target product, remains largely unexplored. Developing such systems could allow for more effective enrichment of high-producing subpopulations and offer a new strategy for improving yield in yeast-based bioproduction.

To investigate whether higher product titres can be achieved in yeast by exploiting non-genotypic cell-to-cell variation, this chapter uses glutathione (GSH) as a model product for the following reasons: (1) *S. cerevisiae* is the main host used for commercial GSH production due to its naturally high intracellular GSH content [268]; (2) GSH is a high-value compound, and previous efforts to improve its production have focused on optimizing culture media, fermentation conditions and genetic engineering strategies such as overexpressing GSH biosynthetic genes [269-275]; (3) GSH biosynthesis in yeast is tightly controlled via a feedback inhibition mechanism [276; 277] (**Figure 6.1**), which provides a means to link GSH level to a growth advantage.

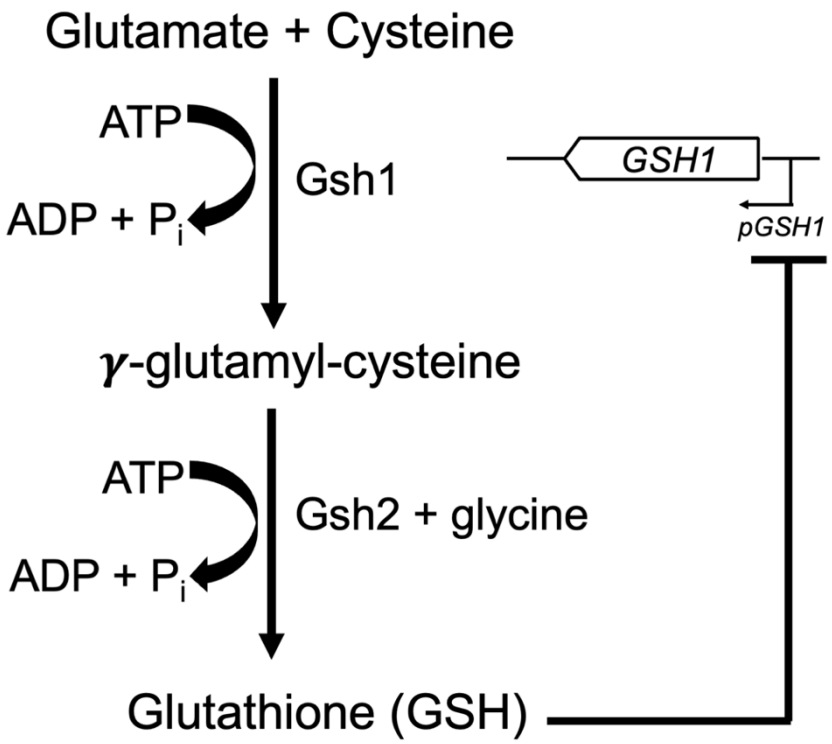


Figure 6.1: GSH biosynthesis pathway in *S. cerevisiae* and feedback regulation of *GSH1* expression. GSH biosynthesis in *S. cerevisiae* involves two ATP-dependent steps. First, cysteine and glutamate are conjugated by the rate-limiting enzyme for GSH biosynthesis: gamma-glutamylcysteine synthetase (Gsh1), which forms the intermediate product gamma-glutamylcysteine. Second, glutathione synthetase (Gsh2) catalyses combination of gamma-glutamylcysteine with glycine to form GSH. A high intracellular level of GSH feedback inhibits expression from the native *GSH1* promoter which controls expression of Gsh1.

6.1.1 Aims

This chapter aimed to construct a regulatory system in *S. cerevisiae* to selectively enrich high-producing cells within a phenotypically heterogeneous, genetically uniform population. GSH was used as a model product, for the reasons outlined above.

The specific objectives were to design and implement a feedback-based expression circuit linking intracellular GSH levels to a growth advantage, and to assess whether this system could enrich high-GSH producers in the population. The heritability of the production state was examined to corroborate whether the observed improvements resulted from non-genotypic, reversible variation.

6.2 Chapter-specific materials and methods

6.2.1 Yeast strains and cultivation conditions

In this chapter, the starting yeast strains used were the haploid *S. cerevisiae* strains BY4741(*MAT***a**; *his3*Δ*1*; *leu2*Δ*0*; *met15*Δ*0*; *ura3*Δ*0*) and BY4742 (*MAT*α; *his3*Δ*1*; *leu2*Δ*0*; *lys2*Δ*0*; *ura3*Δ*0*), and the isogenic diploid strain *S. cerevisiae* BY4743 (*MAT***a**/α; *his3*Δ*1*/*his3*Δ*1*; *leu2*Δ*0*/*leu2*Δ*0*; *lys2*Δ*0*/*LYS2*; *MET15*/*met15*Δ*0*; *ura3*Δ*0*/*ura3*Δ*0*).

Yeasts were maintained and grown either in YPD medium or in YNB medium supplemented as required with amino acids or uracil, 5-FOA (Melford). Wheres specified for particular experiments, D-histidine (Sigma) or reduced GSH (Sigma) were included. Yeast cultures (10 mL) were grown in 50-mL Erlenmeyer flasks from single colonies, at 30 °C, 120 rev. min^{-1} (50 mm orbit).

6.2.2 Quantification of total cellular GSH

Intracellular GSH levels in cells were quantified using a colorimetric assay based on the reduction of 5,5'-dithio-bis-(2-nitrobenzoic acid) (DTNB) (Sigma), according to the protocol from Rahman et al. [278], with minor modifications. Briefly, samples of cells were prepared by diluting exponential phase cells (OD_{600} 1.5–2.0) to an OD_{600} of 1.0 in 1 mL of PBS, and harvested by centrifugation (1000 x *g*, 2 min). The pellets were washed twice with 1 mL PBS to remove residual medium components. Cell pellets were then resuspended in 200 µL of lysis buffer (5% w/v 5-sulfosalicylic acid in 0.1 M potassium phosphate buffer with 5 mM EDTA) together with 425-600 µm-diameter acid-washed glass beads (Sigma) at a buffer/bead ratio of 1:2 by volume. Lysis of cells

was performed with a homogenizer (Precellys Evolution) for six cycles of 20 sec each at 6500 rev. min^{-1}, with 1 min cooling intervals between each cycle. Following lysis, samples were centrifuged at 10,000 x *g* for 10 min at 4 ℃ and the supernatant was used for the DTNB assay. A 200 µL reaction mixture was prepared for each sample in 96-well plates, containing 20 µL of supernatants, 60 µL of 0.66 mg/mL DTNB solution, 60 µL of 0.66 mg/mL β-NADPH (Sigma) and 60 µL of 3.33 units/mL glutathione reductase (Sigma). The reduction of DTNB (colour change) was measured at 412 nm using a Synergy HTX Multi-Mode Reader. Measurements were recorded every 1 min over a 30-min period at 25 °C, with 5 sec of linear shaking (1096 cycles min^{-1}, 1 mm orbit) applied immediately prior to each reading. The rate of change in the absorbance (ΔA_{412nm} min^{-1}) was used to determine the relative concentration of intracellular GSH, based on a standard curve generated from known concentrations of reduced glutathione, as described in Rahman et al. [278].

6.2.3 RNA extraction, cDNA synthesis and Real-Time quantitative PCR (RT-qPCR)

Total RNA was isolated from yeast cells using a TRI Reagent kit (Sigma) according to the manufacturer's instructions. Briefly, samples of cells were prepared by diluting exponential phase cells (OD_{600} 1.5–2.0) to an OD_{600} of 1.0 in 1 mL of PBS, and harvested by centrifugation (1000 x *g*, 2 min). For each sample, 1 mL of TRI reagent was added to cell pellets, together with 425-600 µm-diameter acid-washed glass beads at a reagent/bead ratio of 1:2 by volume. Lysis of cells was performed with the homogenizer for six cycles of 20 sec each at 6500 rev. min^{-1}, with 1-min cooling intervals between each cycle. Phase separation was carried out by addition of 0.2 mL

chloroform to each lysate, followed by vigorous vortexing for 15 sec, incubation at room temperature for 5 min, and centrifugation at 12,000 x *g* for 15 min at 4°C. The colourless upper aqueous phase containing RNA (~400 µL) was transferred to a fresh tube. RNA was precipitated by the addition of 0.5 mL of 2-propanol, followed by 10-min incubation at room temperature. Precipitated RNA was pelleted by centrifugation at 12,000 x *g* for 10 min, washed with 0.5 mL of 75% ethanol, air-dried for 10 min and resuspended in 30 µL nuclease-free water. To remove contaminating DNA, 1 µg of isolated RNA was treated with 1 unit of DNase I (Sigma) at 37 °C for 10 min. The reaction was stopped by the addition of 1 µL of 50 mM EDTA followed by heat inactivation at 70°C for 10 min. RNA concentration and purity were measured by using a spectrophotometer (DeNovix).

cDNA synthesis was performed using the GoScript Reverse Transcription system (Promega), according to the manufacturer's protocol. Briefly, each 20 µL reaction contained 1 µg of DNase-treated RNA, 0.5 µg of oligo$(dT)_{15}$ primer, 4 µL of 5 x reaction buffer, 1.5 µL of $MgCl_2$, 1 µL of PCR nucleotide mix, 1 µL of RNase inhibitor and 1 µL of reverse transcriptase, with nuclease-free water added to the final volume. Reactions were incubated at 25 °C for 5 min (primer annealing), 42 °C for 60 min (reverse transcription), and 70 °C for 15 min (enzyme inactivation). Synthesized cDNA was diluted 1:40 in nuclease-free water prior to RT-qPCR analysis.

RT-qPCR was performed using a QuantStudio 3 Real-Time PCR system (Applied Biosystems) and PowerUP SYBR Green Master Mix (Applied Biosystems). Each 10 µL RT-qPCR reaction contained 5 µL of SYBR Green Master Mix, 500 nM of target gene specific forward and reverse primers, 2 µL of the diluted cDNA as template and

nuclease-free water. Amplification was carried out using the recommended fast cycling mode: 50 °C for 2 min, 95 °C for 2 min, followed by 40 cycles of 95 °C for 5 sec and 60 °C for 1 min. Melt curve analysis was performed at the end of each run to confirm amplification specificity. Each sample was run in technical triplicate. Expression levels of target genes were normalised against the geometric mean of two housekeeping genes, *ACT1* and *UBC6,* which served as internal references. Relative expression levels were calculated using the comparative Ct method [279]. For each target gene, cycle threshold (Ct) values were calibrated to the Ct of the corresponding genes obtained for samples from the control *pCYC1-GAP1* strain and normalised to the geometric mean of the reference genes for relative quantification.

6.3 Results

6.3.1 Construct design for selecting phenotypically high-GSH cells

As cellular GSH is known to feedback inhibit expression from the *GSH1* promoter (*pGSH1*) [276], a feedback-based selection system was designed to enable enrichment of phenotypically high-GSH-producing cells within clonal yeast cultures (**Figure 6.2**). In this system, the expression of the *GSH1 ORF* was placed under the control of the *TEF1* promoter (*pTEF1*), while a counterselectable marker gene was placed under the control of *pGSH1*, which is repressed in response to a high GSH level. The design was intended to allow high-GSH producers to downregulate expression of the counterselectable marker, thereby conferring a selective advantage in the presence of the relevant counterselective agent. In contrast, low-GSH producers would exhibit higher *pGSH1* activity and increased expression of the counterselectable gene, resulting in selective counterselection. The hypothesised net outcome should be a relative enrichment of high-GSH producers within the population. The *pTEF1* promoter was chosen as it bypasses the native GSH-dependent regulation of *GSH1*, being a constitutively active, strong promoter, commonly used in yeast expression systems.

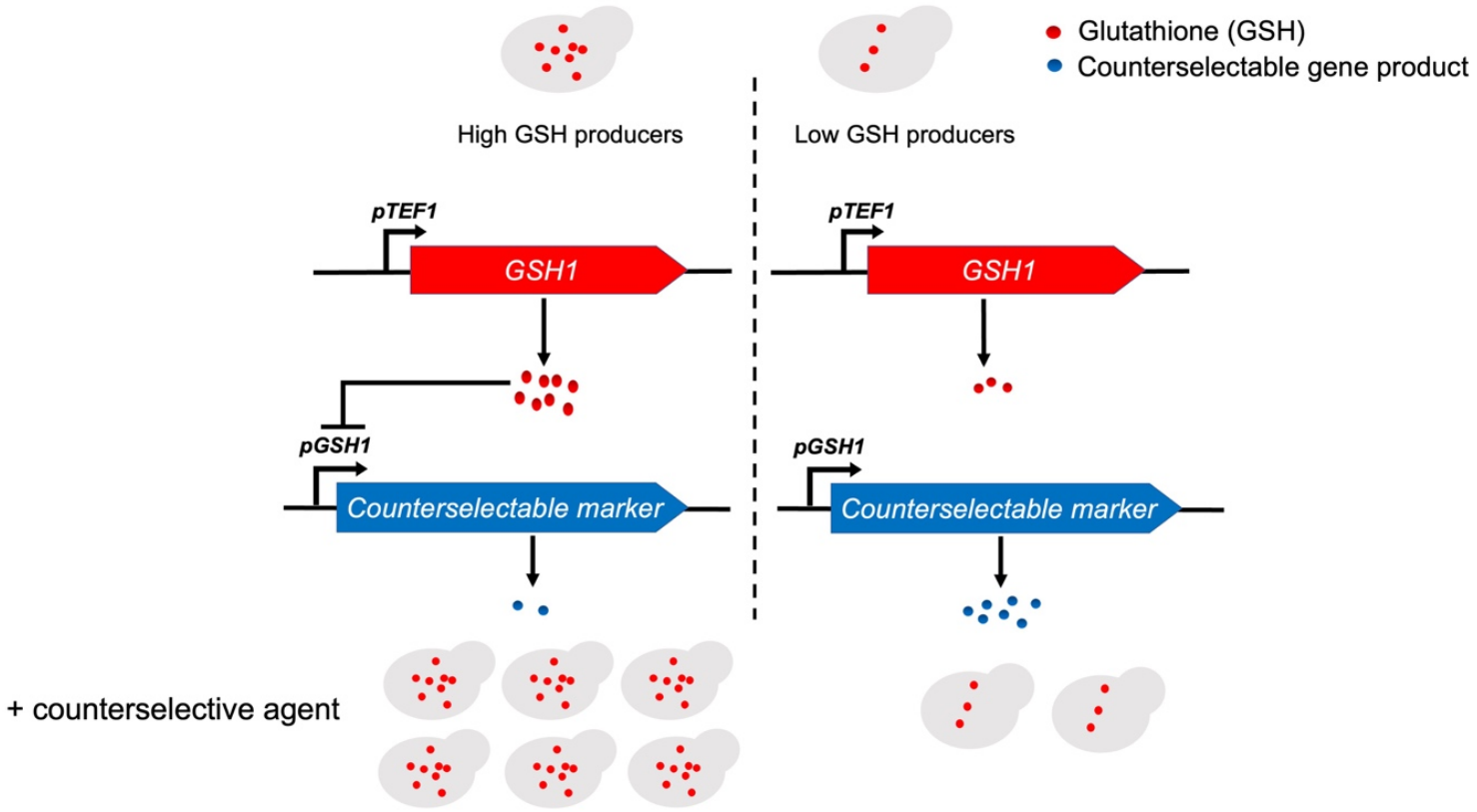


Figure 6.2: Design principle of a feedback-based selection system for enriching phenotypically high-GSH-producing cells. The engineered selection system uses the native feedback regulation of GSH biosynthesis to enrich phenotypically high-GSH-producing cells. The *GSH1* gene is expressed under a constitutive promoter *pTEF1*, while a counterselectable marker is placed under the control of the GSH-responsive *pGSH1*. In high-GSH producers, elevated intracellular GSH level reduces expression of the counterselectable marker from *pGSH1*, allowing these cells to survive under selection, whereas low-GSH producers exhibit higher expression of the counterselectable marker and are eliminated in the presence of the counterselective agent, resulting in a net enrichment of high-GSH-producing cells within the population.

6.3.2 *URA3* is not a suitable counterselectable marker in the system

URA3 was first chosen as the counterselectable marker in the system as described above, based on its widespread application in yeast for counterselection using the drug 5-fluoroorotic acid (5-FOA) [280]. The enzyme encoded by *URA3* converts 5-FOA to the toxic compound 5-fluorouracil, thereby inhibiting the growth of high-*URA3* expressing cells [281]. In the context of the feedback-based selection system, high-*URA3* expressing cells were expected to correspond to low-GSH producers, rendering them sensitive to 5-FOA and allowing selective enrichment of high-GSH producers.

6.3.2.1 Construction of a diploid heterozygous yeast strain incorporating a URA3-based feedback selection system

A diploid, heterozygous yeast strain was constructed in the BY4743 background at the *GSH1* locus, incorporating *URA3* as counter-selectable marker. To do that, the *URA3* ORF was amplified from plasmid pCM190, while the *GFP* coding sequence fused to a *ADH1* terminator sequence (*GFP-tADH1*) and the *HIS3MX6* selection marker were amplified from plasmid pFA6a-GFP(S65T)-His3MX6 [282], the *TEF1* promoter (*pTEF1;* 0–408 bp upstream of the *TEF1 ORF),* and ~300-400-bp homologous sequences upstream (*Up*) and downstream (*Dw*) of the target integration sites for recombination-mediated integration were amplified from BY4741 genomic DNA. Two plasmid constructs were assembled using Gibson assembly (see **Section 2.4.3 for details**). For the first construct, the amplified *URA3* ORF*, GFP-tADH1, Up_gsh1* and *Dw_gsh1* fragments for the *GSH1 ORF* were cloned into pRS315 between the *Sac*I and *Sal*I sites to generate plasmid pRS315-*Up_gsh1-URA3-GFP-tADH1-Dw_gsh1*. For the second construct, the amplified *HIS3MX6, pTEF1, Up_pgsh1* and *Dw_pgsh1* fragments for the *GSH1* promoter region (*pGSH1*) were cloned into pRS315 in a

similar manner to generate plasmid pRS315- *Up_pgsh1-His3MX6-pTEF1-Dw_pgsh1*. The *Up_pgsh1-His3MX6-pTEF1-Dw_pgsh1* fragment was amplified and used to replace the native *pGSH1* in *S. cerevisiae* BY4741 with the strong constitutive promoter *pTEF1* via homologous recombination. Transformants (*pgsh1*Δ::*HIS3MX6-pTEF1*) were selected on YNB agar lacking histidine. The *Up_gsh1-URA3-GFP-tADH1-Dw_gsh1* fragment was amplified and used to replace the *GSH1* ORF in *S. cerevisiae* BY4742 with *URA3-GFP-tADH1,* and transformants (*gsh1*Δ::*URA3-GFP-tADH1*) were selected on YNB agar lacking uracil. The two modified haploid strains were subsequently mated by co-cultivation in YPD broth to generate a heterozygous diploid strain at the *GSH1* locus, with one allele expressing *GSH1* under the control of *pTEF1* and the other carrying *URA3* under *pGSH1* (**Figure 6.3**), and the diploid strain was selected on YNB agar supplemented with 0.1 mg/mL leucine only. Genomic DNA was extracted from candidate colonies, and correct genomic integration events were confirmed by PCR and agarose gel electrophoresis (see **Section 2.3.3 and 2.3.4 for details**).

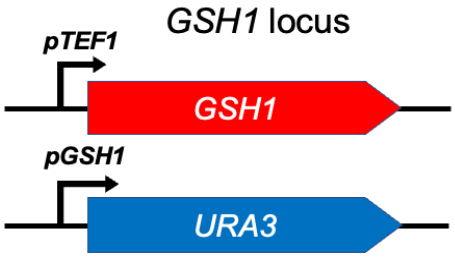


Figure 6.3: Construct design in the heterozygous *pGSH1-URA3* strain. In the heterozygous *pGSH1-URA3* strain, *URA3* is expressed as one allele at the *GSH1* locus and *GSH1* under *pTEF1* control at the other allele.

6.3.2.2 The efficacy of URA3-based feedback selection system in the heterozygous construct strain

To determine the optimal concentration of 5-FOA for the feedback selection, the engineered heterozygous diploid strain was then cultured in YNB medium supplemented with varying concentrations of 5-FOA. A significant increase in cellular GSH level was detected in cultures grown at 100 μg/mL or 150 μg/mL 5-FOA, but not at 50 μg/ml 5-FOA (**Figure 6.4A**). Analysis by flow cytometry revealed that the culture exposed to 150 μg/ml 5-FOA exhibited a high degree of cell clumping (**Figure 6.5A**), which could interfere with OD-based normalization during GSH quantification on a per cell basis. Moreover, culture growth was also reduced at 150 μg/ml 5-FOA (**Figure 6.5B**). Based on these observations, 100 μg/ml 5-FOA was selected as an optimal concentration for further experiments.

To investigate whether the increased GSH level at 100 μg/mL 5-FOA was stable during long-term cultivation of the modified strain, cellular GSH content was monitored every ~10 generations during sequential batch cultivation in flasks at 100 μg/ml 5-FOA (**Figure 6.4B)**. Every ~10 generations, aliquots of overnight culture were transferred to fresh medium at an initial OD_{600} ~0.01 and re-incubated. The GSH level initially increased over the first ~20 generations (i.e. two transfer cycles) but gradually declined over subsequent passages. By approximately 50 generations, (i.e. five transfer cycles), total cellular GSH content had returned to a level close to that of the no addition control cultured in the equivalent condition. These results indicate that although the *URA3*-based feedback selection system appeared initially effective in enhancing total cellular GSH levels, the elevated GSH phenotype was not stably maintained during extended cultivation.

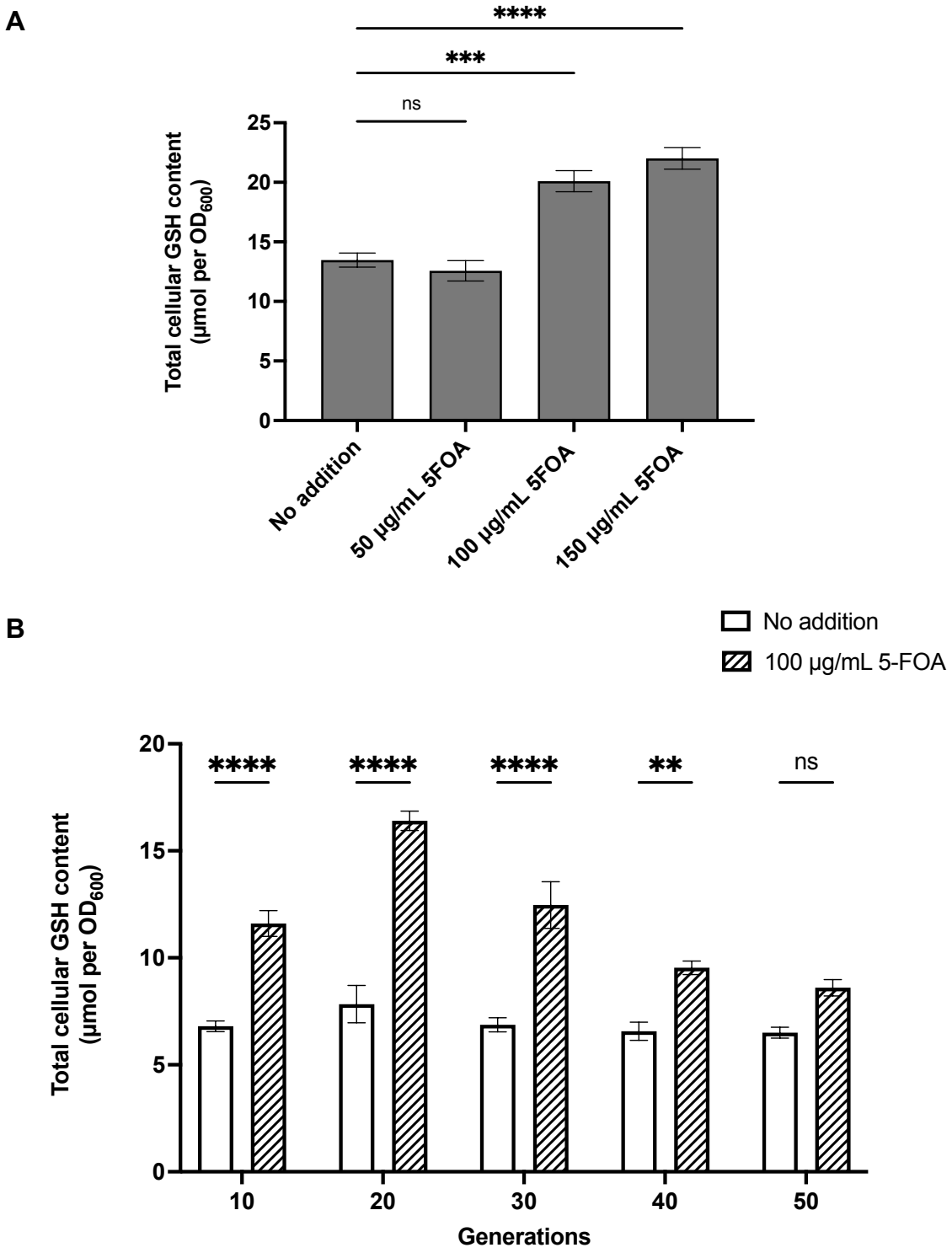


Figure 6.4: Effect of 5-FOA on cellular GSH content of the engineered *S. cerevisiae* containing the heterozygous *GSH1* locus with *URA3* as the counter-selectable marker. (A) GSH contents of the engineered strain shown in **Figure 6.3** after cultivation for 24 h in YNB medium supplemented with 0.02 mg/mL uracil, 0.1

mg/mL leucine and 5-FOA at the indicated concentrations. (B) Long-term cellular GSH level of the engineered strain shown in **Figure 6.3** during sequential batch culture at 100 μg/mL 5-FOA. Cells cultured without 5-FOA were used as a control. Data represent mean values ± SEM of three biological replicates. Statistical analysis: (A) Ordinary one-way ANOVA F (3, 12) = 33.15, $p < 0.0001$, Dunnett's vs no addition (50 μg/mL, ns, $p = 0.7764$; 100 μg/mL, ***, $p = 0.0003$; 150 μg/mL, ****, $p < 0.0001$); (B) two-way ANOVA interaction F(4,20) = 10.04**,** $p = 0.0001$; Šídák (No addition vs 100 μg/mL 5-FOA): 10–30 generations, ****, ($p<0.0001$), 40 generations, **, ($p=0.0067$), 50 generations, ns, ($p=0.0767$).)

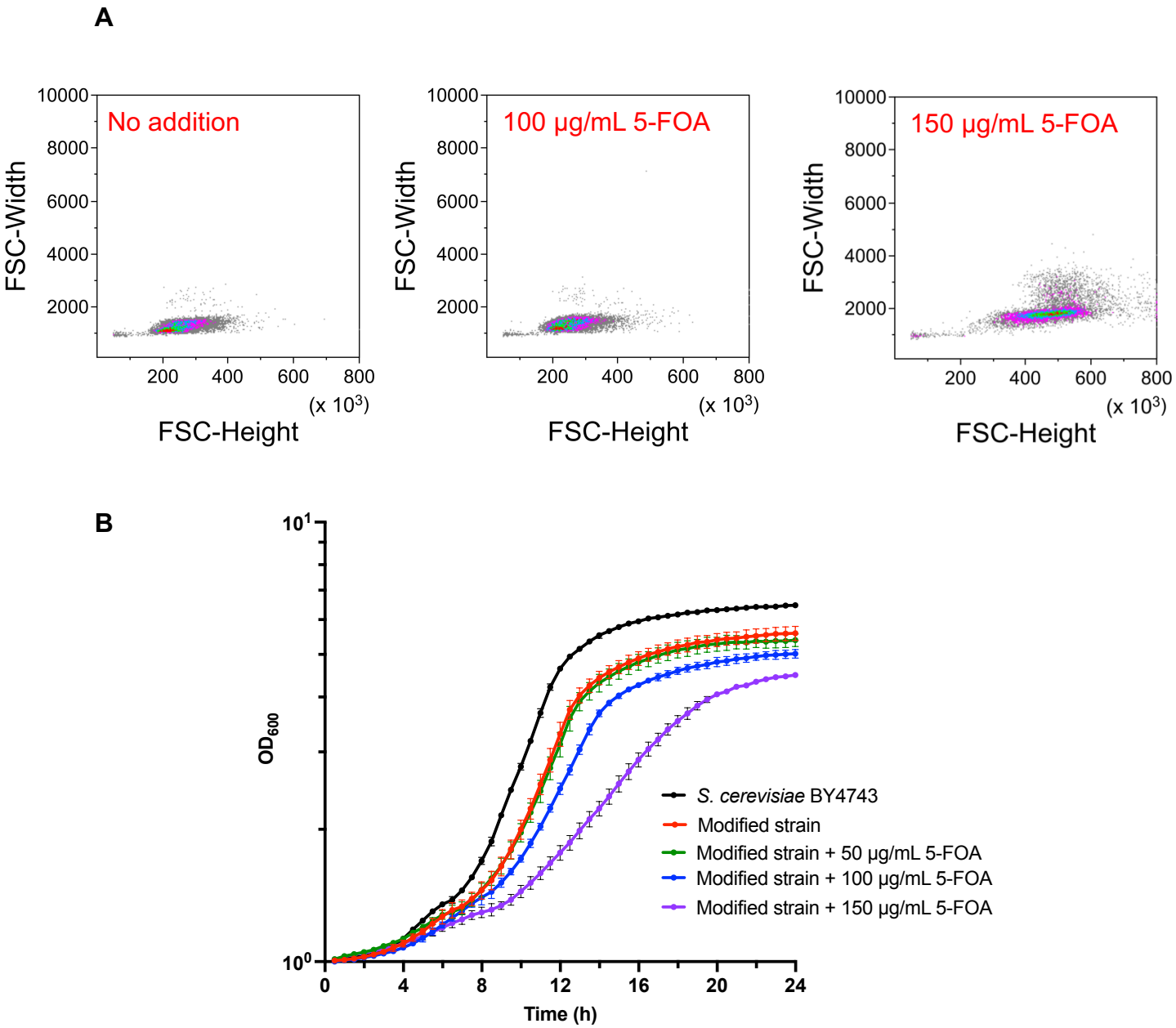


Figure 6.5: Effect of 5-FOA on growth of the engineered *S. cerevisiae* containing the heterozygous *GSH1* locus with *URA3* as counterselectable marker. (A) singlet discrimination by flow cytometry; single cells are in the rectangular gate while cell clumps are outside of the gate according to FSC-height vs FSC-width. Cells cultured without the addition of 5-FOA were used as control. (B) Growth curve of the heterozygous modified strain as shown in **Figure 6.3** with different concentrations of 5-FOA.

6.3.2.3 Loss of URA3 underlies the decline in the GSH level of the heterozygous construct strain

To investigate the cause of the observed decline in GSH levels during prolonged culture, it was hypothesised that the *URA3* allele might have been lost from the genome. To test this, the cultures that had been grown for 50 generations in the presence of 100 µg/mL 5-FOA (**Figure 6.4B**) were plated on YPD agar, and ten individual colonies were randomly selected for genomic DNA extraction and PCR analysis of the *GSH1* locus (**Figure 6.6**). In eight out of ten colonies, the introduced *URA3* gene (~ 2.5 kb) was no longer detectable, and PCR results indicated homozygosity for the *pgsh1*Δ::*HIS3-pTEF1* allele (~ 4.5 kb) at the *GSH1* locus. Moreover, none of the *URA3*-negative colonies were able to grow on minimal medium lacking uracil, confirming functional loss of *URA3*. This loss of heterozygosity could arise by mitotic crossover between homologous chromosomes during cell division [283], resulting in loss of heterozygosity at the *GSH1* locus. Such genetic rearrangements would confer a growth advantage under 5-FOA selection, as cells lacking *URA3* would be resistant to 5-FOA toxicity, thereby leading to the decline in total cellular GSH content.

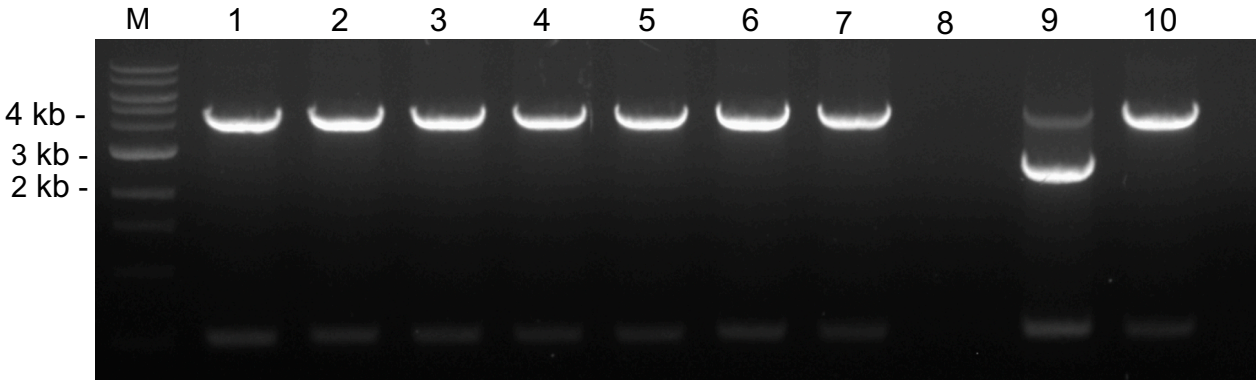


Figure 6.6: PCR analysis of the *GSH1* locus in colonies derived from long-term 5-FOA treated cultures. Cultures of the heterozygous engineered strain grown for approximately 50 generations in the presence of 100 µg/mL 5-FOA (as shown in **Figure 6.4B**) were plated onto YPD agar, and 10 colonies were randomly selected for genomic DNA extraction. PCR was performed using locus-specific primers to assess the integrity of the engineered *GSH1* alleles. Lanes 1-10 show the PCR results for individual colonies. Genotypes unchanged from the starting engineered strain would give two bands, of ~ 2.5 kb (denoting the *gsh1Δ::URA3-GFP* allele) and ~ 4.5 kb (the *pgsh1Δ::HIS3-pTEF1* allele). M: molecular weight marker.

6.3.2.4 Construction of a diploid homozygous yeast strain incorporating a URA3-based feedback selection system

As demonstrated above, the decline in GSH levels observed in the heterozygous construct could be due to the loss of the *URA3* counterselectable marker. It was hypothesised that a homozygous configuration would provide greater genetic stability. In such a strain, mitotic crossover between homologous chromosomes would not result in the loss of *URA3* marker, as both loci would carry the same construct. To test

this, a homozygous diploid strain was engineered with modified alleles at both the *GSH1* and *URA3* loci. To construct the homozygous diploid strain, the *pGSH1* sequence (0–585 bp upstream of the *GSH1* ORF), and *Up_ura3* and *Dw_ura3* sequences of the *URA3* ORF were amplified from *S. cerevisiae* BY4741 genomic DNA. These elements, together with the previous *GFP-tADH1* fragment were assembled into the pRS315 vector using Gibson assembly to generate the plasmid pRS315-*Up_ura3-pGSH1-URA3-GFP-tADH1-Dw_ura3*. The *Up_pgsh1-His3MX6-pTEF1-Dw_pgsh1* and *Up_ura3-pGSH1-URA3-GFP-tADH1-Dw_ura3* fragments were then amplified from the relevant pRS315 constructs. Each cassette was used to transform both the BY4741 and BY4742 strains, to replace *pGSH1* with *pTEF1* and insert *URA3-GFP-tADH1* in the *URA3* locus, creating *pgsh1*Δ::*His3MX6-pTEF1; ura3*Δ::*pGSH1-URA3-GFP-tADH1* in both the BY4741 and BY4742 backgrounds. The resulting two haploid strains were then mated to generate a diploid strain, designated *pGSH1-URA3,* which was homozygous at both the *GSH1* and *URA3* loci, with *pTEF1*-driven *GSH1* expression and *pGSH1*-driven *URA3* expression present on both alleles (**Figure 6.7A**). A diploid homozygous control strain was also constructed in which *URA3* was expressed under its native promoter *pURA3* (**Figure 6.7B**). This *pURA3-URA3* strain was generated in a similar manner as described above, using an integration cassette containing *pURA3* amplified from pCM190, instead of *pGSH1*.

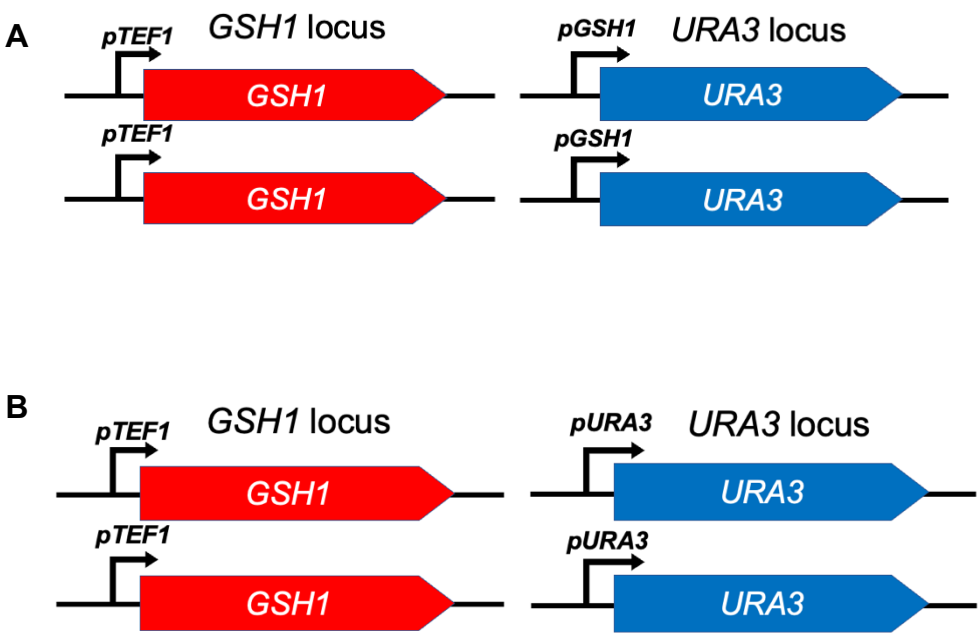


Figure 6.7: Construct designs in the homozygous *pGSH1-URA3* strain and the *pURA3-URA3* control strain. (A) Construct design in the homozygous diploid pGSH1-URA3 strain with GSH1 under pTEF1 control at the GSH1 locus and URA3 under pGSH1 control at the URA3 locus. (B) Construct design in the homozygous pURA3-URA3 control strain with GSH1 under pTEF1 control at both alleles of the GSH1 locus and URA3 under its native promoter pURA3 at both alleles of the URA3 locus.

6.3.2.5 Enhanced genetic stability of the URA3-based feedback selection system in a homozygous construct

To assess the stability of the GSH-enhancing phenotype, the homozygous *pGSH1-URA3* strain was cultured in sequential batch in the presence of 100 µg/mL 5-FOA, and total cellular GSH levels were monitored every ~ 10 generations, following the same procedure as described earlier for the heterozygous strain. An approximate 50% increase in cellular GSH content was observed in the cultures supplemented with 5-FOA, and this increase was stably maintained throughout the cultivation period (**Figure 6.8**). This indicated improved phenotypic stability conferred by the homozygous arrangement.

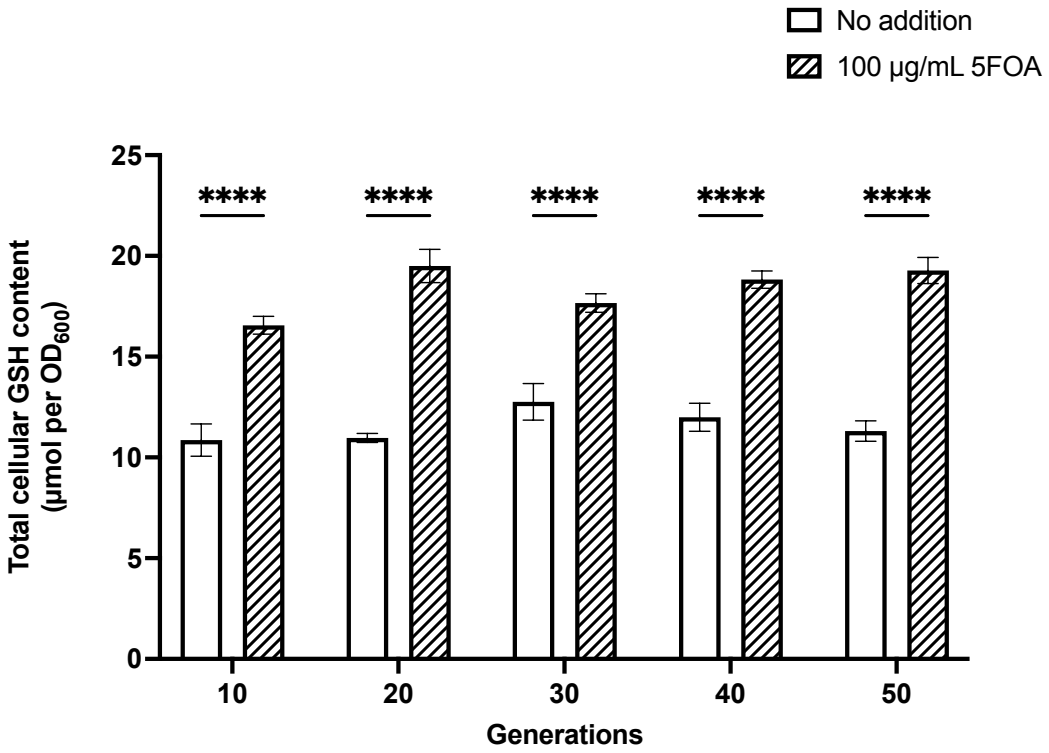


Figure 6.8: Effect of 5-FOA on cellular GSH content of the homozygous engineered *S. cerevisiae* with *URA3* as the counter-selectable marker. Long-term cellular GSH level of the homozygous engineered strain shown in **Figure 6.7A** during sequential batch culture with 100 µg/mL 5-FOA. Cells cultured without 5-FOA were used as controls. Data represent mean values ± SEM of three biological replicates.

Statistical significance (two-way ANOVA, treatment × time): $F(4,30) = 2.901$, $p = 0.0384$; Šídák (No addition vs 100 μg/mL 5-FOA): 10–50 generations, ****, $p < 0.0001$.

6.3.2.6 The increase in GSH level was a generic response to 5-FOA rather than selective enrichment

Although the homozygous *pGSH1-URA3* strain above demonstrated stable elevation of cellular GSH levels when cultured in the presence of 100 μg/mL 5-FOA, it remained uncertain whether this increase was due to selective enrichment of phenotypically high-GSH-producing cells or instead reflected a generic cellular response to 5-FOA. To investigate this, a control strain designated *pURA3-URA3* was constructed (**Figure 6.7B**). Unexpectedly, this control strain also exhibited a ~50% increase in total cellular GSH content when cultured with 5-FOA, comparable to the increase observed in the *pGSH1-URA3* strain (**Figure 6.9**). In addition, expression of *URA3* in the *pGSH1-URA3* strain was examined via flow cytometry, according to levels of GFP-tagged Ura3. If the feedback-based selection system was functioning as intended, selection for high-GSH producers should lead to inhibition of Ura3-GFP expression via GSH-mediated repression of *pGSH1*. However, a significant increase in Ura3-GFP fluorescence was observed upon addition of 5-FOA (**Figure 6.10**), indicating that *pGSH1* activity was upregulated rather than repressed under that condition. These findings suggest that the observed GSH elevation in both the *pGSH1-URA3* and *pURA3-URA3* strains was due to a general stress or metabolic response to 5-FOA, rather than selective enrichment of phenotypically high-GSH-producing cells. Moreover, the unexpected activation of *pGSH1* in the presence of 5-FOA further undermines the specificity and reliability of the *URA3*-based feedback selection system. It is therefore concluded that *URA3* is not a suitable counterselectable marker in this system.

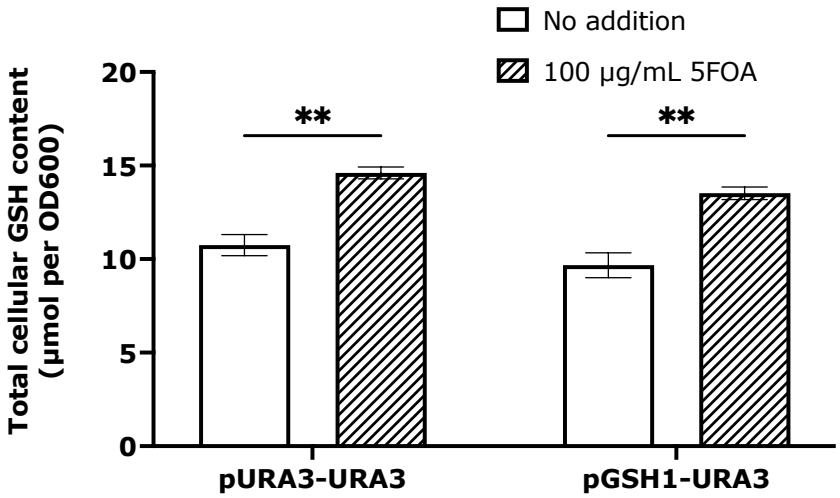


Figure 6.9: Cellular GSH level of the control homozygous *pURA3-URA3* strain under 5-FOA selection. Cellular GSH levels in the control homozygous pURA3-URA3 strain shown in **Figure 6.7A**, compared to those in the *pGSH1-URA3* strain with 100 µg/mL 5-FOA. Cultures without additions were used as controls. Bars represent the mean ± SEM of three biological replicates. Statistical analysis (two-way ANOVA, treatment effect): $F(1, 8) = 60.82$, $p<0.0001$; Šídák (No addition vs 100 µg/mL 5-FOA): $p=0.0011$, **, for both strains.

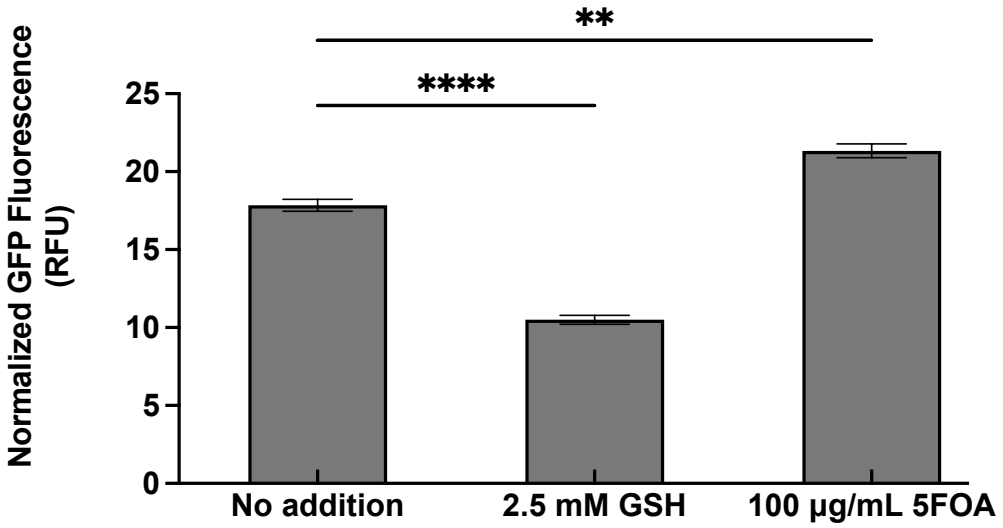


Figure 6.10: Ura3-GFP expression in the homozygous *pGSH1-URA3* strain.

GFP fluorescence was used to assess Ura3-GFP expression in the modified homozygous *pGSH1-URA3* strain. Cells cultivated without additions served as a negative control, while supplementation with 2.5 mM reduced GSH was used as a positive control (addition of GSH should give a reduction of Ura3-GFP expression). GFP fluorescence of cells was measured by flow cytometry and normalised to cell size (FSC-A). Background fluorescence was subtracted using measurements from wild type *S. cerevisiae* BY4743 strain. Bars represent the mean of three biological replicates. Statistical significance (Ordinary one-way ANOVA): $F(2, 6) = 217.0$, $p < 0.0001$; pairwise differences are indicated as $p < 0.01$ (**) and $p < 0.0001$ (****).

6.3.3 *GAP1* as an alternative counterselectable marker in the system

6.3.3.1 Construction of a diploid homozygous yeast strain incorporating a GAP1-based feedback selection system

As described in the previous section, *URA3* was found to be unsuitable as a counter-selectable marker for the feedback selection system. Therefore, *GAP1*, a native yeast

gene encoding a general amino acid permease [284], was explored as an alternative. Cells expressing Gap1 can be selectively inhibited by D-histidine (D-HIS), which is toxic and transported into cells exclusively via Gap1 [285]. A homozygous diploid strain was engineered, designated *pGSH1-GAP1*, in which the native *GAP1* promoter was replaced with the GSH-responsive *pGSH1* promoter at both *GAP1* alleles, and the *GSH1* was under the control of the constitutive promoter *pTEF1* (**Figure 6.11A**). To do this, a *KanMX* fragment was amplified from pFA6A-KanMX6 [282], and *Up_gap1* and *Dw_gap1* fragments for the *pGAP1* region (~ 300-400 bp upstream of the *GAP1 ORF*) were amplified from BY4741 genomic DNA. These fragments plus *pGSH1* were assembled into the pRS315 vector using Gibson assembly to generate pRS315-*Up_gap1-KanMX-pGSH1-Dw_gap1*. The two cassettes *Up_pgsh1-His3MX6-pTEF1-Dw_pgsh1* (described in **Section 6.3.2.1**) and *Up_gap1-KanMX-pGSH1-Dw_gap1* were amplified from their respective plasmids and used to replace *pGSH1* with *pTEF1* and *pGAP1* with *pGSH1* in both of the *S. cerevisiae* BY4741 and BY4742 strains (giving genotype *pgsh1*Δ::*His3MX6-pTEF1; pgap1*Δ::*KanMX-pGSH1*). Transformants were selected on YNB agar lacking histidine for replacement of *pGSH1* with *pTEF1* and on YPD agar containing 200 µg/mL G418 for replacement of *pGAP1* with *pGSH1*. The resulting haploid strains were then mated as above by co-cultivation in YPD broth. The diploid homozygous strain *pGSH1-GAP1* was selected on YNB agar supplemented with 0.02 mg/mL uracil, 0.1 mg/mL leucine. A diploid homozygous control *pCYC1-GAP1* strain was also constructed using the same approach as described above, using an integration cassette containing the *CYC1* promoter (*pCYC1*) (0–300 bp upstream of the *CYC1* ORF) amplified from BY4741 genomic DNA, instead of *pGSH1* (**Figure 6.11B**).

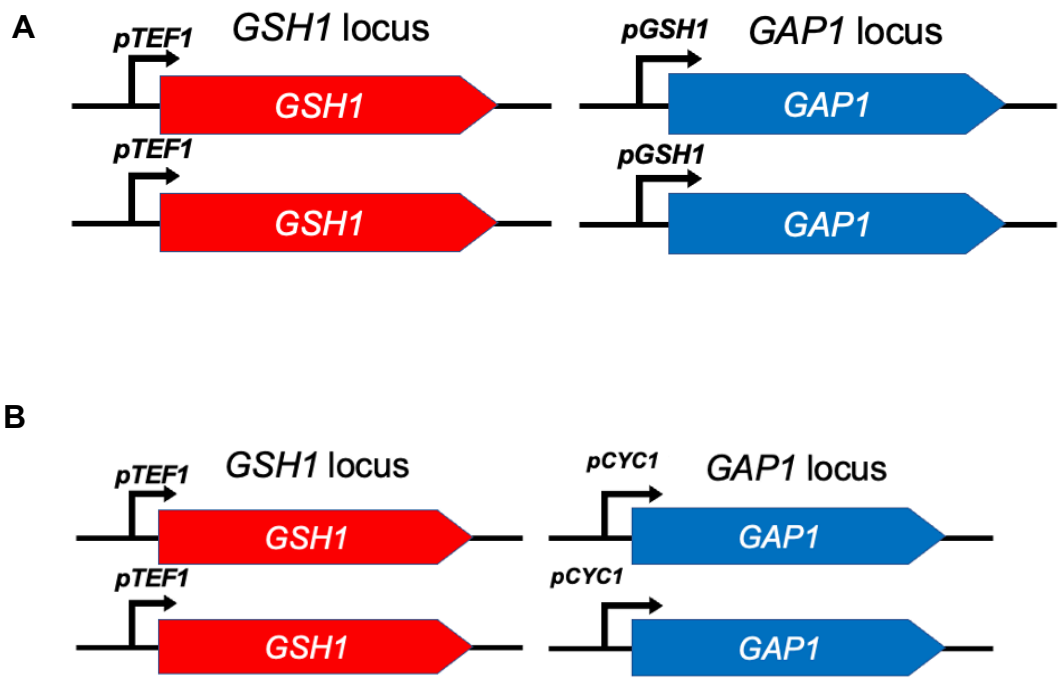


Figure 6.11: Construct designs in the homozygous *pGSH1-GAP1* strain and the *pCYC1-GAP1* control strain. (A) Construct design in the homozygous diploid *pGSH1-GAP1* strain with *GSH1* under *pTEF1* control at the *GSH1* locus and *GAP1* under GSH-responsive *pGSH1* control at the *GAP1* locus. The modified strain designated *pGSH1-GAP1*. (B) Construct design in the homozygous *pCYC1-GAP1* control strain with *GSH1* under *pTEF1* control at both alleles of the *GSH1* locus and *GAP1* under constitutive *pCYC*1 promoter at both alleles of the *GAP1* locus. The modified strain designated *pCYC1-GAP1*.

6.3.3.2 The efficacy of GAP1-based feedback selection system in the homozygous construct strain

GAP1 expression is naturally repressed by rich nitrogen sources such as ammonium sulphate and glutamine and is induced in the presence of a poor nitrogen source [286]. A poor nitrogen source can result in a low growth rate, which itself might affect the rate of GSH synthesis non-specifically. However, in the *pGSH1-GAP1* construct, *GAP1* expression is no longer regulated by nitrogen quality, but instead by intracellular GSH levels via *pGSH1*. This design allows the use of standard YNB medium containing ammonium sulphate.

To determine the optimal concentration of D-HIS for the feedback selection system, the *pGSH1-GAP1* strain was cultured in the presence of a range of D-HIS concentrations in YNB medium. The total cellular GSH content was significantly increased at 600 μg/mL D-HIS (**Figure 6.12A**). However, increasing the D-HIS concentration to 800 μg/mL returned the GSH level close to the no-addition control, and both growth rate and growth yield of cultures were significantly decreased at this higher concentration (**Figure 6.12A and Figure 6.13**). Therefore, 600 μg/mL D-HIS was selected as an optimal concentration for subsequent experiments. To investigate whether this increase in GSH level was stable during long-term cultivation, the *pGSH1-GAP1* strain was cultivated over approximately 60 generations under 600 μg/mL D-HIS selection. The elevated cellular GSH content was sustained throughout long-term culture, although the size of the effect did decrease a little over time (**Figure 6.12B**).

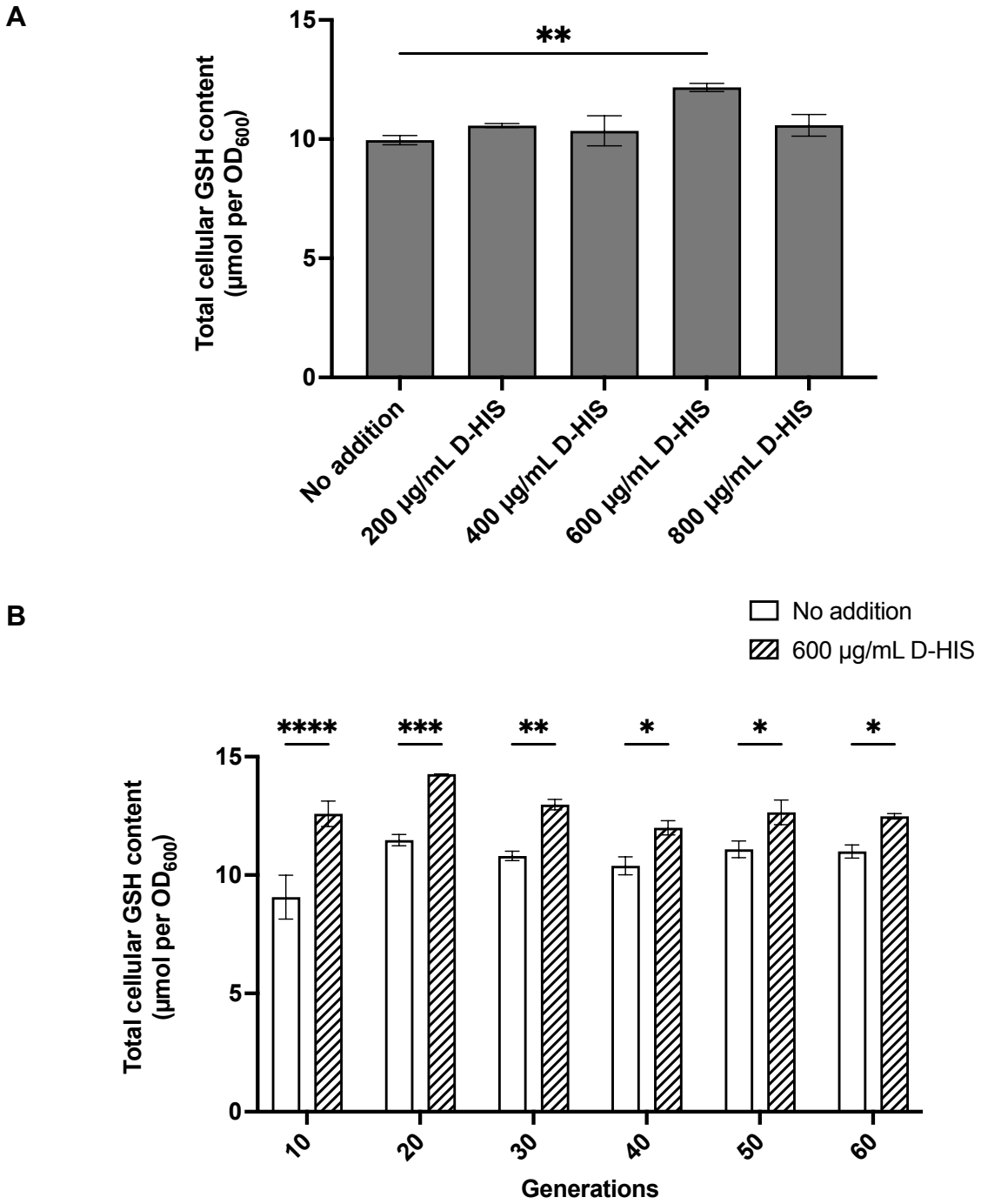


Figure 6.12: Effect of D-HIS on cellular GSH content of the homozygous engineered *S. cerevisiae* with *GAP1* as the counter-selectable marker. (A) Total cellular GSH contents of the pGSH1-GAP1 strain after cultivation for 24 h in YNB medium supplemented with D-HIS at the indicated concentrations. (B) Long-term cellular GSH level of the homozygous engineered strain shown in **Figure 6.7A** during sequential batch culture with 600 µg/mL D-HIS. Cells cultured without D-HIS were

used as a control. Bars represent the mean ± SEM of three biological replicates. Statistical analysis: (A) One-way ANOVA: F (4, 10) = 5.277, p=0.0151; Dunnett's vs no addition: 600 µg/mL, **, (p=0.0057); others, ns. (B) Two-way ANOVA: F (5, 12) = 3.261, p = 0.0434; Šídák (No addition vs 600 µg/mL): 10–20 generations, p≤0.0003, 30 generations, p=0.0027, 40–60 generations, p≤0.0399.

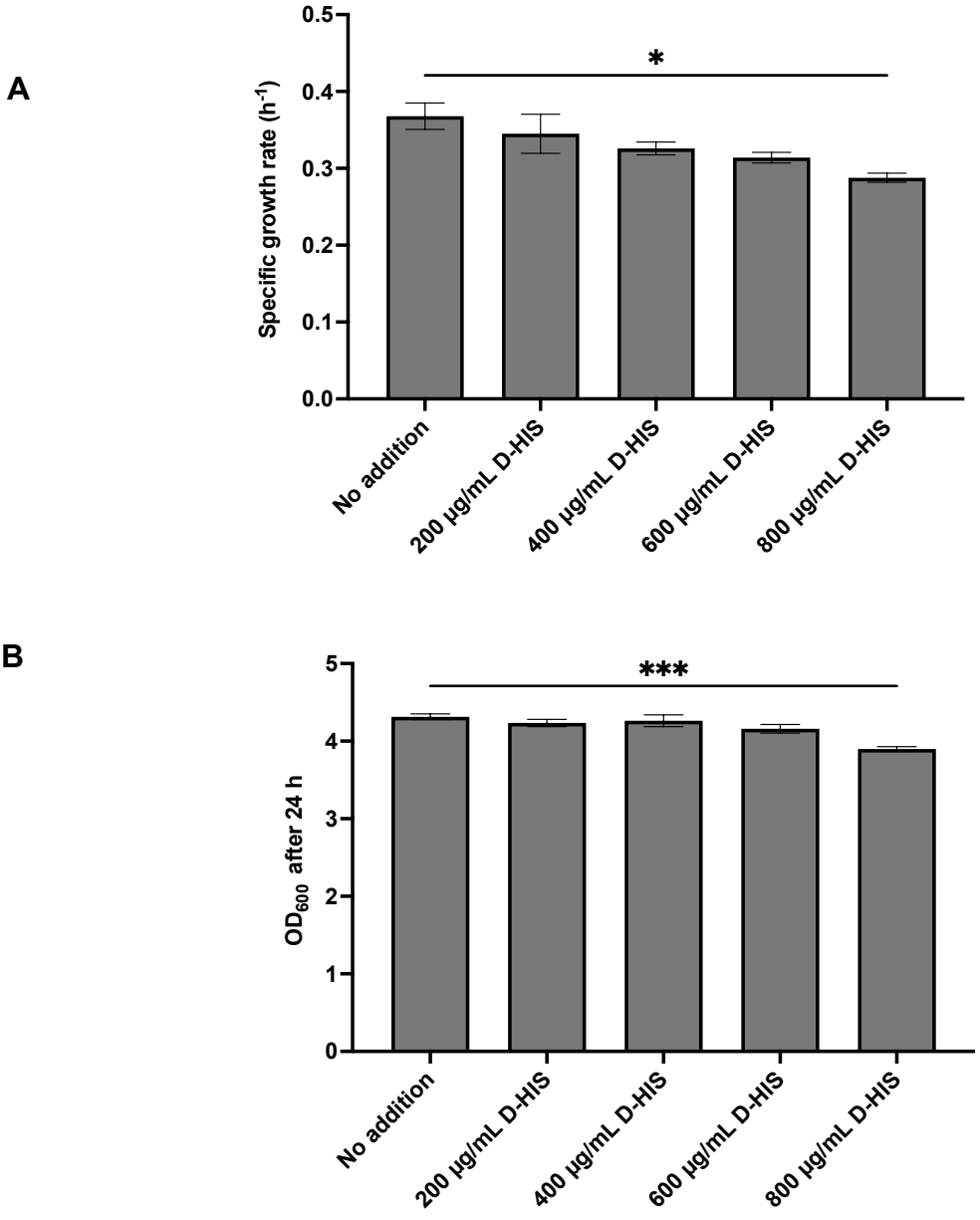


Figure 6.13: Effect of D-HIS on growth of the homozygous engineered *S. cerevisiae* with *GAP1* as counterselectable marker. Cells were cultivated for 24 h in YNB medium supplemented with D-HIS as the indicated concentrations. (A) Specific growth rate of the *pGSH1-GAP1* strain. (B) OD_{600} of the *pGSH1-GAP1* strain. Bars represent the mean ± SEM of three biological replicates. Statistical analysis (ordinary one-way ANOVA): (A) F (4, 10) = 4.216, p = 0.0296; (B) F (4, 10) = 10.04, p = 0.0016. Post hoc significance is indicated by * and **.

6.3.3.3 The increase in GSH level was not a general response to D-HIS

To exclude the possibility that the observed increase in GSH level in the *pGSH1-GAP1* strain was a general response to D-HIS exposure rather than specific consequence of feedback selection, a control strain designated *pCYC1-GAP1* was constructed (**Figure 6.11B**). (The native *pGAP1* promoter was not suitable for this comparison because its activity is repressed by the ammonium sulphate used as the nitrogen source in the culture medium, which would prevent D-HIS uptake and confound the analysis). The *pCYC1* promoter was selected for this because it supports moderate and constitutive expression and based on transcriptomic data in YPD culture, it drives *GAP1* expression at a level similar to *pGSH1* [287]. Therefore, both *pCYC1-GAP1* and *pGSH1-GAP1* strains were expected to exhibit comparable levels of *GAP1* expression and consequently similar uptake of D-HIS under the tested conditions. Unlike the *pGSH1-GAP1* strain, which showed elevated GSH levels in response to D-HIS exposure, the *pCYC1-GAP1* control strain exhibited a decrease in GSH content when cultured with D-HIS (**Figure 6.14**). This inverse response rules out the possibility that the GSH increase observed in the *pGSH1-GAP1* strain was due to a general stress

response or other non-specific effect of D-HIS. Instead, the result supports a conclusion that the elevated GSH level in the *pGSH1-GAP1* strain when cultured with D-HIS reflects the specific outcome of the feedback-driven selection.

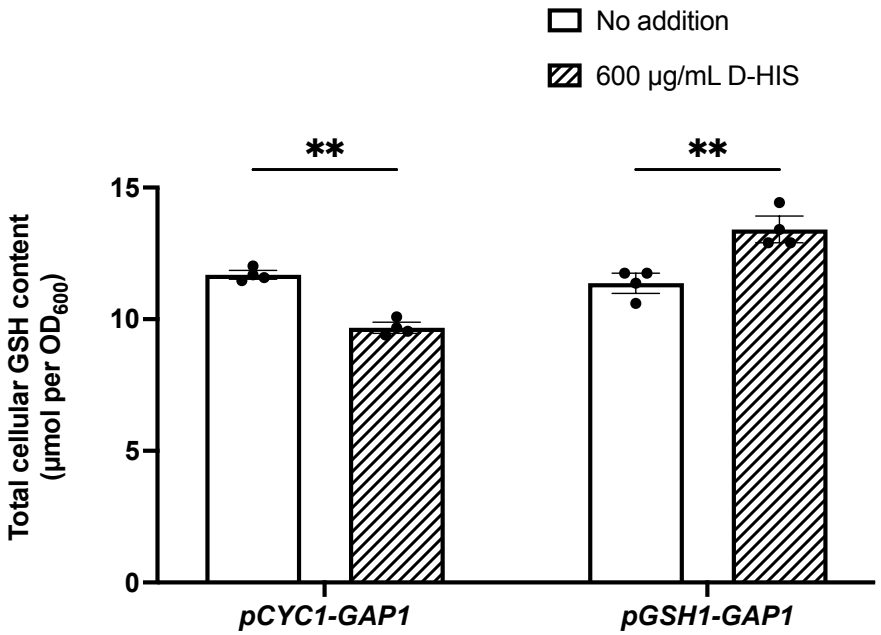


Figure 6.14: Cellular GSH level of the control homozygous *pCYC1-URA3* strain under 5-FOA selection. Cellular GSH levels in the control homozygous *pCYC1-GAP1* strain shown in **Figure 6.11B** compared to those in the *pGSH1-GAP1* strain, with 600 µg/mL D-HIS. Cultures without additions were used as controls. Bars represent the mean ± SEM of three biological replicates. Statistical analysis (two-way ANOVA): F (1, 8) = 34.31, p = 0.0004; Šídák (No addition vs 600 µg/mL D-HIS): p = 0.0068 (pCYC1-GAP1) and p = 0.0062 (pGSH1-GAP1) (**).

6.3.3.4 The high-GSH-producing phenotype was non-heritable

To exclude the possibility that the high-GSH-producing phenotype observed after long-term cultivation with D-HIS might result from the emergence and fixation of stable genetic mutants, the heritability of the high-GSH-producing phenotype was assessed.

The *pGSH1-GAP1* strain, which had been cultured for approximately 60 generations in the presence of 600 µg/mL D-HIS (as shown in **Figure 6.12B**), was transferred to D-HIS-free YNB medium and cultured for 24 h. Cells from this culture were then subcultured into fresh YNB medium with or without D-HIS. It was found that elevated GSH levels were only re-established when cells were re-exposed to D-HIS (**Figure 6.15**). In the absence of D-HIS, total cellular GSH content returned to the untreated control level. Given that the phenotype was observable within as few as 10 generations of initial D-HIS exposure (**Figure 6.12B**), it is unlikely that the response resulted from the emergence and selection of spontaneous mutations. Instead, the high-GSH-producing phenotype is better explained by a reversible, non-heritable response mediated by the engineered feedback system.

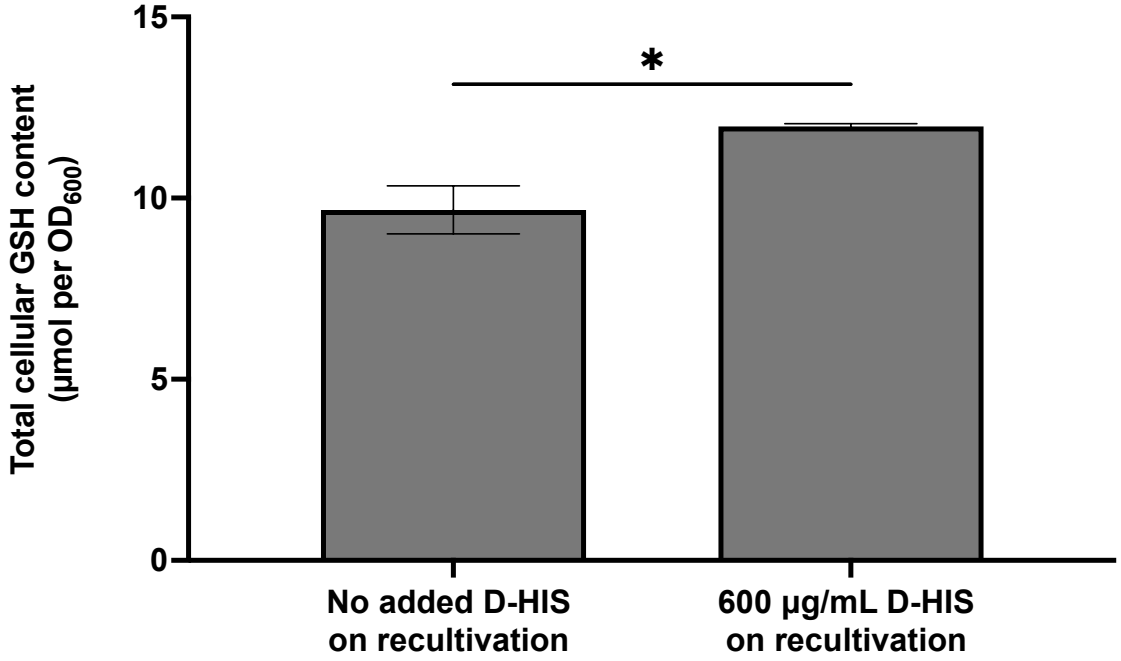


Figure 6.15: Lack of heritability of the high-GSH-producing phenotype after removal of selection pressure. The *pGSH1-GAP1* strain previously cultured for 60

generations in the presence of 600 µg/mL D-HIS (as shown in **Figure 6.12B**) was transferred to YNB medium without D-HIS and grown for 24 h. Cells were then subcultured from OD_{600} ~ 0.1 into fresh YNB medium with or without D-HIS and grown for an additional 24 h. Total cellular GSH levels were measured to evaluate the stability of the high-GSH phenotype. Bars represent the mean ± SEM of three biological replicates. Statistical significance (two-tailed unpaired t-test): $t(4) = 3.446$, $p = 0.0262$ (*).

6.3.3.5 RT-qPCR analysis reveals selection-dependent expression patterns of GSH-related genes

To further support the observations and conclusions, transcript levels of the counterselectable marker gene *GAP1* and the GSH biosynthesis genes *GSH1* and *GSH2* were quantified by RT-qPCR in the *pGSH1-GAP1* strain. As expected from the feedback control design, where *GAP1* is under the control of the GSH-responsive *pGSH1* promoter, a marked reduction in *GAP1* transcript level was observed in cells cultivated with D-HIS (**Figure 6.16A**). This result is consistent with selection for cells producing higher intracellular GSH levels, which repress *pGSH1* activity and thereby lower *GAP1* expression. Moreover, expression of both *GSH1* and *GSH2* was found to be greater in the D-HIS-selected population relative to the untreated control, supporting the conclusion that the selected population indeed exhibited enhanced GSH biosynthesis capacity (**Figure 6.16B and C**). A positive control, where the *pGSH1-GAP1* strain was treated with 1 mM extracellular reduced GSH, also showed a similar repression of *GAP1* (**Figure 6.16A**). Interestingly, this condition also included an increase in the *GSH1* transcript level, suggesting the possibility of *pTEF1* regulation

by relatively high (≥1 mM) extracellular GSH. In contrast, the *pCYC1-GAP1* negative-control strain, where *GAP1* expression is driven constitutively by *pCYC1*, showed no significant changes in *GAP1, GSH1* or *GSH2* transcript levels in response to D-HIS treatment (**Figure 6.16**). The evidence collectively indicated that the high-GSH phenotype observed under D-HIS selection in the *pGSH1-GAP1* strain resulted from the selective enrichment of phenotypically high-GSH-producing cells.

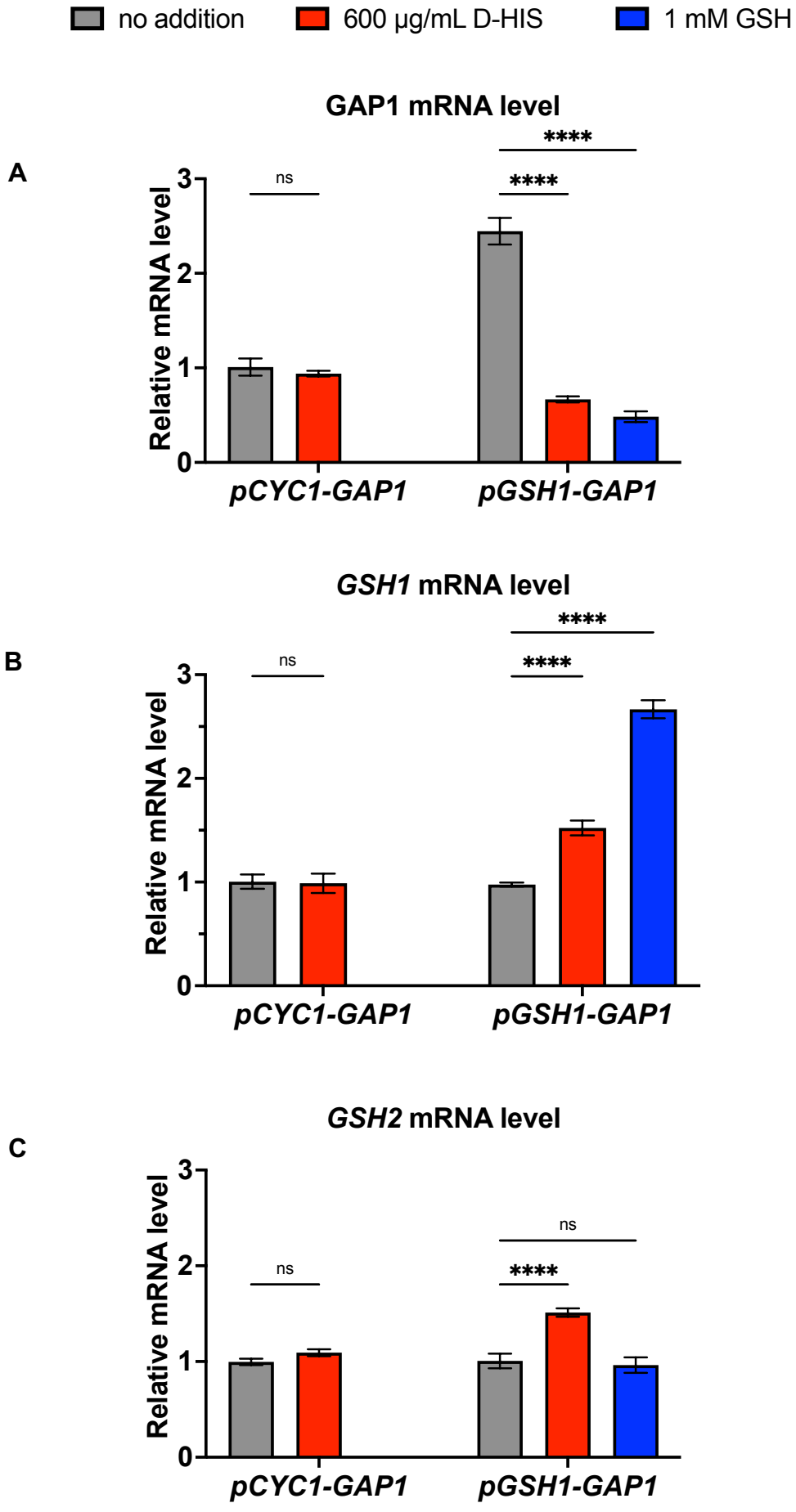

no addition
600 μg/mL D-HIS
1 mM GSH
A
GAP1 mRNA level
Relative mRNA level
ns
****
****
pCYC1-GAP1
pGSH1-GAP1
B
GSH1 mRNA level
Relative mRNA level
ns
****
****
pCYC1-GAP1
pGSH1-GAP1
C
GSH2 mRNA level
Relative mRNA level
ns
ns
****
pCYC1-GAP1
pGSH1-GAP1

Figure 6.16: Effect of D-HIS on *GAP1*, *GSH1*, and *GSH2* mRNA levels in engineered *pGSH1-GAP1* strain. The *pGSH1-GAP1* strain was cultured for 24 h in YNB medium with or without 600 µg/ml D-HIS. As controls, the *pCYC1-GAP1* strain was included as a negative control, and the *pGSH1-GAP1* strain cultured in YNB medium supplemented with 1 mM extracellular reduced GSH was included as a positive control. Relative mRNA levels of *GAP1* (A), *GSH1* (B), *GSH2* (C) in RNA extracts were quantified by RT-qPCR. Bars represent means ± SEM of three biological replicates. Statistical analysis (two-way ANOVA): (A) $F(2, 12) = 71.34$; (B) $F(2, 12) = 282.0$; (C) $F(2, 12) = 55.21$. $p < 0.0001$ (****), ns: not significant.

6.3.4 Cysteine supplementation enhanced the effect of feedback selection on GSH production

It is well established that intracellular GSH levels in yeast are influenced by nutrient composition, particularly the availability of amino acid precursors such as cysteine and glycine [273-275]. As the experiments described above were conducted in minimal YNB medium, where cysteine availability might be a limitation, it was hypothesised that medium supplementation with cysteine could facilitate increased GSH levels in the *pGSH1-GAP1* strain under D-HIS selection. To test this, the *pGSH1-GAP1* strain was cultivated for 24 h in YNB medium with or without 0.1 mM cysteine and with or without 600 µg/mL D-HIS. In the absence of cysteine, D-HIS treatment led to an ~18% increase in total cellular GSH content. However, when 0.1 mM was included in the medium, the addition of D-HIS resulted in a ~ 50% increase in GSH levels (**Figure 6.17**). These results suggest that medium optimization through cysteine supplementation could enhance the performance of the feedback selection system,

possibly by broadening the phenotypic heterogeneity of intracellular GSH levels within the population and thereby enabling more effective enrichment of high-producing subpopulations under D-HIS selection.

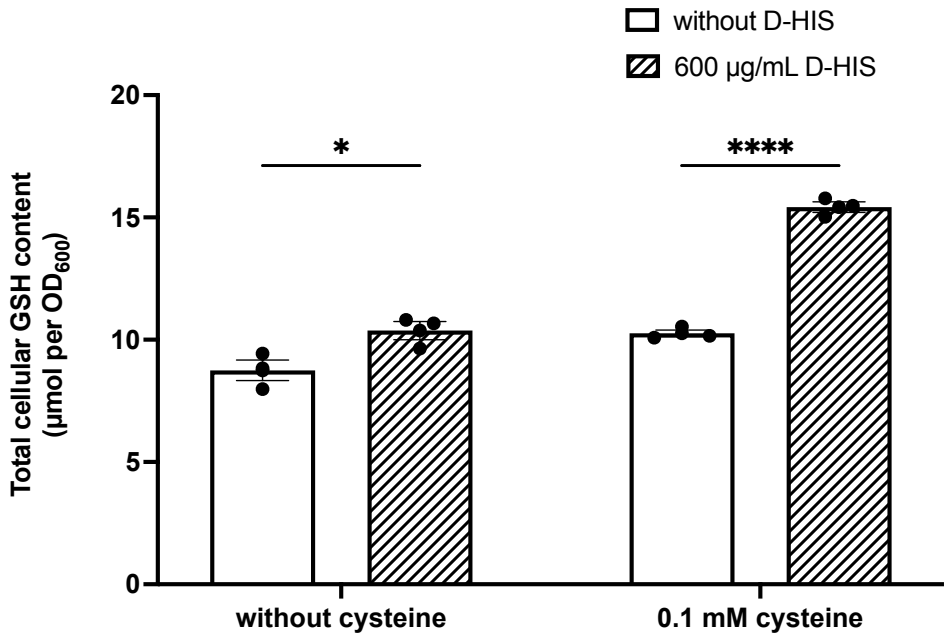


Figure 6.17: Effect of cysteine supplementation on cellular GSH levels in the engineered *pGSH1-GAP1* strain. Total cellular GSH content was measured in the *pGSH1-GAP1* strain after 24 h cultivation in YNB medium under four conditions: (i) without D-HIS and cysteine, (ii) with 600 µg/mL D-HIS only, (iii) with 0.1 mM cysteine only, (iv) with both 600 µg/mL D-HIS and 0.1 mM cysteine. Bars represent the mean ± SEM of three biological replicates. Statistical analysis (two-way ANOVA, D-HIS × cysteine): $F(1, 8) = 32.76$, $p = 0.0004$; Šídák (No addition vs 600 µg/mL D-HIS): without cysteine, *, ($p=0.0118$); 0.1 mM cysteine, ****, ($p<0.0001$).

6.4 Discussion

6.4.1 Leveraging phenotypic heterogeneity for production enhancement

Cell-to-cell phenotypic variation is a common feature in any genetically uniform cell population. In industrial biomanufacturing, this variability can present a significant challenge. High-producing cells often carry a heavier metabolic load, which compromises their growth and competitive fitness. As a result, low-producing cells may have a growth advantage, reducing productivity. Previous efforts in yeast have focused on mitigating this decline by stabilizing production for longer through synthetic circuits that respond to pathway intermediates or product precursors [79; 267]. In contrast, the approach described in this chapter takes advantage of the existing phenotypic diversity to enhance biosynthesis by applying selective pressure to favour the growth of high-producing subpopulations. While this concept has been demonstrated in *E. coli*, using product-responsive circuits to enrich for non-genetic high producers [145], its application in yeast remains unexplored. GSH was used as an exemplar yeast product, because its intracellular levels are known to vary between cells, and its well-characterised regulatory pathway provides a suitable framework for the development of a product-responsive selection system [213; 276]. The heterogeneity in GSH content within a yeast cell population is driven at least partly by metabolic oscillations and transcriptional regulatory mechanisms [213].

6.4.2 Impact of marker choice on feedback selection system performance

The effectiveness of the feedback selection system for enriching phenotypically high-producing cells may critically rely on an appropriate counterselectable marker. This was clearly demonstrated in this chapter, where the *URA3*-based system failed to

achieve selective enrichment. Although increased GSH levels were observed in strains exposed to 5-FOA, this effect was also present in a control strain expressing *URA3* under its native promote *pURA3*, indicating that the response was not specific to phenotypic selection. Instead, it likely reflected a general cellular response to 5-FOA, undermining the intended function of the system. Notably, *URA3* expression from the heterozygous construct was found to be genetically unstable under 5-FOA selection. A rapid loss of heterozygosity was observed, likely due to mitotic recombination, a common genome rearrangement mechanism in yeast associated with DNA repair and adaptation [288]. The resulting homozygous *ura3* mutants, rendered insensitive to 5-FOA, should rapidly dominate the population, effectively eliminating the selective pressure. These findings highlight the importance of genetic stability when designing selection-based control circuits. In contrast, *GAP1* proved to be a more suitable marker for phenotype-based selection. When placed under the control of the GSH-responsive *pGSH1* promoter, *GAP1* enabled effective enrichment of high-GSH-producing subpopulations using D-HIS as a counterselective agent. In the control strain expressing *GAP1* under the control of a constitutive *pCYC1* promoter, the cellular GSH level was reduced to some extent in the presence of D-HIS. This suggests that D-HIS exposure might non-specifically decrease the cellular GSH concentration. As a result, the actual increase in GSH observed with the feedback system might be underestimated, and even greater improvement could be possible with alternative counterselection markers that exert less background response on the cells.

The 18% increase in cellular GSH content achieved through the *GAP1*-based feedback selection system, and the 50% increase observed when also cysteine-

supplemented, represent a promising foundation for future optimisation efforts in yeast-based production. In the case of GSH, the counterselection system could complement or further enhance the increased GSH content that is attainable through gene knockout or overexpression approaches [213; 289]. Additionally, nutrient availability plays an important role in determining intracellular GSH levels. While cysteine is a well-known limiting precursor, other nutrients, such as glutamic acid, glycine and peptone, also contribute to GSH biosynthesis and may influence production outcomes [274; 290]. Therefore, further optimisation of the growth medium could improve the overall effectiveness of the system.

6.4.3 Practical considerations and potential of the feedback selection system

Previous studies have shown that production lifetimes of products such as vanillin-*β*-glucoside and N-acetylglucosamine in yeast can be extended by regulating the population heterogeneity via control circuits responding to the target product intermediates, so the product yield was increased when measured in the long term [79; 267]. In this chapter, a different approach was demonstrated: enhancement of native product biosynthesis, specifically GSH, through counterselection against low-producing cell subpopulations, using a circuit responding directly to the target product. Notably, this enhancement was observable not only over extended cultivation but also in the short term (i.e. within 10 generations). In experiments with *E. coli,* a similar strategy gave approximately three-fold enhancement of production of free fatty acid and tyrosine [145], compared to the ~18-50 % increase in GSH observed in this chapter with yeast. Several factors may explain this difference. First, the extent of cell-to-cell variation in product levels (i.e. on which selection pressure can be exerted

through these approaches) may be different between host species or pathways. Second, the *E. coli* strains into which the product-responsive circuits selecting nongenetic high producers were introduced had previously been engineered for overproduction of fatty acid and tyrosine, through gene knockouts, overexpression and/or heterologous gene expression [145]. In contrast, the feedback system described in this chapter was introduced into a wild-type laboratory *S. cerevisiae* background. As suggested above, it is reasonable to expect that future integration of this selection system into engineered yeast strains optimised for GSH production could lead to greater yield than those observed here.

One limitation of the feedback control system developed in this chapter is that it relies on a native product that is subject to feedback inhibition, such as GSH, for which a responsive regulatory framework already exists. This restricts its direct application to a limited range of naturally regulated metabolites. To extend its use, particularly for heterologous proteins or non-native products, one potential solution is to implement a co-expression system that allows coordinated expression of a gene of interest (GOI) and a feedback-regulated gene, such as *GSH1*, from the same promoter. In this setup, cells with higher expression of *GSH1* and consequently higher GSH levels, would also express the GOI at elevated levels, allowing simultaneous selection for both. Such coordinated expression to produce separate proteins can be achieved, for example, by incorporating internal ribosome entry sites (IRES) or a self-cleaving peptides [291]. A previous study demonstrated that using a 2A peptide sequence from the equine rhinitis B virus to link the red fluorescence protein and GFP gene sequences yielded comparable expression levels to their monocistronic expression [292]. Applying similar

designs could enable coupling of GSH-based selection to the production of unrelated target proteins, thereby extending the use of this selection system.

Another practical consideration is that the current counter-selectable markers for yeast are limited and the counterselective agents, such as 5-FOA and D-HIS, can be relatively expensive [280], which may constrain industrial-scale applications. While costs may come down or new markers or selective agents may be developed in the future, it is important to recognise that the GSH-based feedback system served as an effective proof of principle, and other unexplored regulatory systems in yeast could potentially offer more robust or scalable strategies for enrichment of high-producing individuals for commercial exploitation.

6.4.4 Conclusion

This chapter demonstrated that phenotypic heterogeneity in yeast can be exploited to improve product yield through a feedback-based selection system. By linking intracellular GSH levels to growth advantage, high-producing subpopulations were selectively enriched from an isogenic population. While the *URA3*-based system was hindered by genetic instability and non-specific stress responses, *GAP1* served as a more suitable counterselectable marker in this system. Overall, this chapter provides proof-of-principle that endogenous regulatory mechanisms can be used to enrich phenotypically high producers in yeast. While demonstrated with GSH, the strategy may be extended to other products using co-expression designs and alternative regulatory elements in the future.

7 SUMMARY AND CONCLUDING REMARKS

7.1 Key themes and discussion

This thesis demonstrates that production by yeast populations during biomanufacturing processes can be subject to effects of genetic mutation, host-specific factors and non-genotypic heterogeneity between cells. These processes may interact dynamically under production stress, their interplay shaping both the stability and performance of production strains. Across four experimental chapters, the work showed that production stability depends on product type and host background (Chapter 3), that mutations in production plasmids are dominant cause of productivity decline (Chapter 4), that stress-associated phenotypic heterogeneity further reduce productivity (Chapter 5), and that heterogeneity can be harnessed through feedback-based genetic systems to enhance yields (Chapter 6). The themes emerging from this thesis provide insights and considerations for production strain design and optimisation.

7.1.1 The fitness-productivity trade-off

One key theme is the fitness-productivity trade-off in high-producing strains. While high production levels are essential for industrial productivity, they impose a significant metabolic burden that drives selection for variants with reduced expression. These adaptive changes can occur through either genetic mutation or non-genotypic phenotype shifts that reduce production and restore cellular fitness. Although there was a general trend in which higher productivity was associated with faster production decline, not all strains responded to production stress in the same way. Even among strains with similar initial production levels, marked differences were observed in their

rates of decline. This indicates that host genomic background influences the capacity to tolerate stress and modulates the trajectory of evolutionary adaptation. These findings provide valuable insights into how strain selection and engineering can be optimised to achieve more robust and sustained bioproduction. For example, genetically diverse strain libraries can be used for screening to identify high-performing, stable production hosts, and stress tolerance pathways in production stains can be engineered to relieve the burden and prolong productive lifespan. Such approaches have been validated in recent studies where the exploitation of natural yeast strain diversity yielded improved production [162; 293], and where stress response engineering in yeast maximised titres under industrial conditions [294]. The strain-focused strategy may be further expanded by exploring interspecies yeast diversity, which offers an even broader pool of genotypes and phenotypes for selection.

7.1.2 Mechanisms of production decline

Another important theme is mechanisms resulting in the production decline. Two distinct but complementary mechanisms of production decline were uncovered in this thesis: genetic instability of the expression system and stress-associated phenotypic heterogeneity. The first mechanism involved the accumulation of mutations in the expression plasmid. These mutations clustered in specific regions, including the *MET17* promoter region, the *FRT* sites and the heterologous coding sequence. Such hotspots likely represent structurally or functionally sensitive loci that are under strong selective pressure. Their repeated appearance across evolved populations suggests that these loci represent functionally sensitive regions where mutations are strongly selected under production conditions. Such regions should therefore be priority targets for stabilisation in future plasmid design. For example, synthetic promoters and codon-

optimised coding sequences should be refined with sequence context in mind, as mutation rates have been shown to be sequence dependent [295]. Although the 2-micron plasmid based expression system has been shown to be less stable than genome-integration for *S. cerevisiae* bioproduction [296], certain mutations identified here would still be relevant when considering genome-integration based bioproduction. The second mechanism contributing to the production decline was stress-related phenotypic heterogeneity. High-producing cell subpopulations tended to exhibit elevated stress levels than low-producing cell subpopulations, and cultures with higher average productivity also displayed greater heterogeneity in stress responses compared to lower-producing cultures. As higher stress levels are usually associated with reduced growth rates due to higher physiological cost [297], cells in high-producing state are likely to be outcompeted by faster-growing cells in lower-producing states. Because production phenotypes within an isogenic population are transient, with individual cells able to switch between high and low production over short timescales, intracellular stress emerges as a key factor driving phenotypic divergence and production instability. These findings also provided potential strategies for improving productivity, such as engineering stress tolerance pathways or optimising cultivation conditions to minimise stress levels in production populations [294; 298; 299].

7.1.3 Use of phenotypic heterogeneity

While phenotypic heterogeneity could be viewed as a source of inefficiency due to cell-cell variability in a desired phenotype, this thesis demonstrates that it can be harnessed to improve production. Coupling the target product to a selectable fitness trait enabled enrichment of high-producing cells. The use of a feedback-regulated

selection system allowed the endogenous metabolic state of the cell to guide population composition, favouring the persistence of high producers. Although the increase in production achieved through this system in this thesis was moderate, approximately 18% under standard YNB medium and up to 50% with a single medium supplement, it provides a clear proof-of principle that phenotypic heterogeneity can be reshaped to optimise yeast strains for bioproduction. Importantly, combining such feedback control with complementary strategies, such as selecting robust starting strains from genetically diverse libraries and engineering stress tolerance pathways to help alleviate intracellular burden, could offer promising approaches for developing superior production strains.

7.2 Future work

Based on the findings and insights of this thesis, several directions emerge for future research to deepen the understanding of production instability and to enhance the long-term performance of yeast-based cell factories.

7.2.1 Cell cycle contributions to phenotypic heterogeneity

A contributor to the phenotypic heterogeneity observed in recombinant protein output and stress readouts may be variation in cell cycle stage. In *S. cerevisiae*, many cellular processes relevant to protein production, including transcriptional programmes, proteostasis demand and cell size, change systematically across the cell cycle, and a substantial fraction of genes show periodic expression [300]. Therefore, some of the cell-to-cell variability may reflect asynchronous cell cycle distribution rather than stable phenotypic subpopulations. A direct way to dissect this effect in future work would be to introduce cell cycle synchronisation as an experimental control. For example,

pheromone treatment using alpha-factor can synchronise MATa cells by arresting them in G1, after which production and stress reporters can be quantified at defined time windows to map output onto cell cycle progression [301]. If synchronisation reduces variance at comparable mean expression, this would support a significant cell cycle contribution to the measured heterogeneity.

7.2.2 Systems-level understanding of production instability

Although this thesis identified key mechanisms underlying production decline, a more comprehensive systems-level understanding is needed to uncover how genetic background, cellular physiology and regulatory or metabolic pathways interact during long-term cultivation. Further investigation is required to clarify how genotype influences phenotype at the molecular level and to uncover stress response mechanisms that may contribute to phenotypic heterogeneity. These insights would provide a stronger foundation for rational engineering of production strains. Multi-omics approaches (e.g., transcriptomics, proteomics, metabolomics) offer useful tools to track global cellular changes during long-term production and, potentially, help discover early markers of production instability. Specifically, performing multi-omics analysis on the distinct non-, low- and high-producing subpopulations identified at the initial production stage could provide valuable insights into the molecular basis of this heterogeneity, e.g., heterogeneously expressed regulatory or metabolic components or metabolic imbalances that distinguish high-producing cells from non- and low-producing cells, ultimately guiding strategies for improving both stability and yield.

7.2.3 Discovery and engineering of robust host genotypes

The present demonstration that strain background plays an important role in production stability supports a future focus on identifying the genetic determinants that enable certain strains to sustain high-yield production. Genome-wide association studies and QTL mapping offer potential tools for this purpose. By comparing high-producing, stable strains with high-producing but unstable ones, it may be possible to identify specific single-nucleotide polymorphisms or other genetic differences that correlate with enhanced production stability. Once identified, such candidate loci can be subjected to functional validation using genetic engineering. Confirmed targets could then be introduced into production strains to improve their robustness. This genotype-guided approach offers a more predictive and efficient alternative to empirical screening, supporting the rational development of next-generation industrial yeast strains with improved performance and reliability.

7.2.4 Application beyond the model system

Finally, future research should extend the findings of this thesis beyond the models used here, i.e., the test proteins, *S. cerevisiae* and standard laboratory conditions. While *S. cerevisiae* is a well-established host for biomanufacturing, it may not always be the most suitable chassis for every application. Exploring non-conventional yeast species or alternative microbial platforms could uncover hosts with more favourable characteristics in terms of protein secretion, stress tolerance, genetic stability or metabolic efficiency. Furthermore, it is important to assess whether the strategies developed in this work are effective at larger scales. All yeast strains in this thesis were cultured under controlled laboratory conditions in 48-well plates or in 50-mL shake flasks. Therefore, validation of strain performance and system robustness in

small bioreactor or pilot-scale fermentation systems will be essential to ensure that the observed benefits may translate to industrial bioprocess conditions. This will also allow the evaluation of production stability under more realistic conditions, such as fed-batch fermentation, where factors like oxygen limitation, shear stress or nutrient gradients may influence strain behaviour [73; 74; 302]. Bridging the gap between bench-scale proof-of-principle and industrial deployment will be a critical step toward establishing practical, scalable platforms for sustainable biomanufacturing.

REFERENCES

1. Zhang, Y. P., Sun, J. and Ma, Y. (2017). Biomanufacturing: history and perspective. *Journal of Industrial Microbiology & Biotechnology,* **44**, 773-784.

2. Demain, A. L., Vandamme, E. J., Collins, J. and Buchholz, K. (2017). History of industrial biotechnology. *Industrial Biotechnology*. 1-84.

3. McGovern, P. E., Zhang, J., Tang, J., Zhang, Z., Hall, G. R., Moreau, R. A., Nunez, A., Butrym, E. D., Richards, M. P., Wang, C. S., Cheng, G., Zhao, Z. and Wang, C. (2004). Fermented beverages of pre- and proto-historic China. *Proceedings of the National Academy of Sciences of the United States of America,* **101**, 17593-17598.

4. Sano, C. (2009). History of glutamate production. *American Journal of Clinical Nutrition,* **90**, 728S-732S.

5. BCC Research (2023). *Commercial amino acids* [online]. Available at: https://www.bccresearch.com/market-research/biotechnology/commercial-amino-acids.html (Accessed 10 April 2025).

6. Sanchez, S., Rodríguez-Sanoja, R., Ramos, A. and Demain, A. L. (2018). Our microbes not only produce antibiotics, they also overproduce amino acids. *The Journal of Antibiotics,* **71**, 26-36.

7. Spiteller, P. (2015). Chemical ecology of fungi. *Natural Product Reports,* **32**, 971-993.

8. Hagedorn, S. and Kaphammer, B. (1994). Microbial biocatalysis in the generation of flavor and fragrance chemicals. *Annual Review of Microbiology,* **48**, 773-800.

9. Berlec, A. and Strukelj, B. (2013). Current state and recent advances in biopharmaceutical production in *Escherichia coli,* yeasts and mammalian cells. *Journal of Industrial Microbiology & Biotechnology,* **40**, 257-274.

10. Fortune Business Insights (2025). *Biopharmaceuticals market size, share & industry analysis, by type, enzymes, vaccines, hormones, by application, by distribution channel, and reginal forecast, 2024-2032* [online]. Available at: https://www.fortunebusinessinsights.com/biopharmaceuticals-market-106928. (Accessed 10 April 2025).

11. Hong, K. K. and Nielsen, J. (2012). Metabolic engineering of *Saccharomyces cerevisiae*: a key cell factory platform for future biorefineries. *Cellular and Molecular Life Sciences,* **69**, 2671-2690.

12. Khan, A. H., Tye, G. J. and Noordin, R. (2020). CRISPR-Cas9 genome editing tool for the production of industrial biopharmaceuticals. *Molecular Biotechnology,* **62**, 401-411.

13. Zeng, W., Guo, L., Xu, S., Chen, J. and Zhou, J. (2020). High-throughput screening technology in industrial biotechnology. *Trends in Biotechnology,* **38**, 888-906.

14. Goh, J. B. and Ng, S. K. (2018). Impact of host cell line choice on glycan profile. *Critical Reviews in Biotechnology,* **38**, 851-867.

15. Shapira, P., Matthews, N. E., Cizauskas, C. A., Aurand, E. R., Friedman, D. C., Layton, D. S., Maxon, M. E., Palmer, M. J. and Stamford, L. (2022). Building a bottom-up bioeconomy. *Issues in Science and Technology,* **38**, 78-83.

16. Yang, Y. and Sha, M. (2019). A beginner's guide to bioprocess modes–batch, fed-batch, and continuous fermentation. *Eppendorf Incorporated Application Note,* **408**, 1-16.

17. Behl, M., Thakar, S., Ghai, H., Sakhuja, D. and Bhatt, A. K. (2023). Chapter 21 - Fundamentals of fermentation technology. *Basic Biotechniques for Bioprocess and Bioentrepreneurship*: Academic Press. 313-328.

18. Kim, M.-J., Park, S.-L., Kim, H.-J., Sung, B. H., Sohn, J.-H. and Bae, J.-H. (2024). Functional expression of recombinant insulins in *Saccharomyces cerevisiae*. *Microbial Cell Factories,* **23**, 302.

19. Mohd Azhar, S. H., Abdulla, R., Jambo, S. A., Marbawi, H., Gansau, J. A., Mohd Faik, A. A. and Rodrigues, K. F. (2017). Yeasts in sustainable bioethanol production: A review. *Biochemistry and Biophysics Reports,* **10**, 52-61.

20. Anastassiadis, S., Morgunov, I. G., Kamzolova, S. V. and Finogenova, T. V. (2008). Citric acid production patent review. *Recent Patents on Biotechnology,* **2**, 107-123.

21. Abdel-Rahman, M. A., Tashiro, Y. and Sonomoto, K. (2013). Recent advances in lactic acid production by microbial fermentation processes. *Biotechnology Advances,* **31**, 877-902.

22. Mu, X. and Zhang, F. (2023). Diverse mechanisms of bioproduction heterogeneity in fermentation and their control strategies. *Journal of Industrial Microbiology and Biotechnology,* **50**.

23. Ding, Q. and Ye, C. (2023). Microbial cell factories based on filamentous bacteria, yeasts, and fungi. *Microbial Cell Factories,* **22**, 20.

24. Wang, Y., Liu, L., Jin, Z. and Zhang, D. (2021). Microbial cell factories for green production of vitamins. *Frontiers in Bioengineering and Biotechnology,* **9**, 661562.

25. Baeshen, M. N., Al-Hejin, A. M., Bora, R. S., Ahmed, M. M., Ramadan, H. A., Saini, K. S., Baeshen, N. A. and Redwan, E. M. (2015). Production of biopharmaceuticals in *E. coli*: current scenario and future perspectives. *Journal of Microbiology and Biotechnology,* **25**, 953-962.

26. Wosten, H. A. B. (2019). Filamentous fungi for the production of enzymes, chemicals and materials. *Current Opinion in Biotechnology,* **59**, 65-70.

27. Ferrer-Miralles, N., Domingo-Espin, J., Corchero, J. L., Vazquez, E. and Villaverde, A. (2009). Microbial factories for recombinant pharmaceuticals. *Microbial Cell Factories,* **8**, 17.

28. Mattanovich, D., Sauer, M. and Gasser, B. (2014). Yeast biotechnology: teaching the old dog new tricks. *Microbial Cell Factories,* **13**, 34.

29. Rettenbacher, L. A., Arauzo-Aguilera, K., Buscajoni, L., Castillo-Corujo, A., Ferrero-Bordera, B., Kostopoulou, A., Moran-Torres, R., Núñez-Nepomuceno, D., Öktem, A., Palma, A., Pisent, B., Puricelli, M., Schilling, T., Tungekar, A. A., Walgraeve, J., Humphreys, D., von der Haar, T., Gasser, B., Mattanovich, D., Ruddock, L. and van Dijl, J. M. (2022). Microbial protein cell factories fight back? *Trends in Biotechnology,* **40**, 576-590.

30. Liu, T., Xu, X., Liu, Y., Li, J., Du, G., Lv, X. and Liu, L. (2022). Engineered microbial cell factories for sustainable production of L-lactic acid: A critical review. *Fermentation,* **8**, 279.

31. Berovic, M. and Legisa, M. (2007). Citric acid production. *Biotechnology annual review,* **13**, 303-343.

32. Waché, Y. and Dijon, A. (2013). Microbial production of food flavours. *Microbial production of food ingredients, enzymes and nutraceuticals*: Woodhead Publishing. 175-193.

33. Zabermawi, N. M., Alsulaimany, F. A. S., El-Saadony, M. T. and El-Tarabily, K. A. (2022). New eco-friendly trends to produce biofuel and bioenergy from microorganisms: An updated review. *Saudi Journal of Biological Sciences,* 10.1016/j.sjbs.2022.02.024.

34. Tang, X.-L., Yu, J.-W., Geng, Y.-H., Wang, J.-R., Zheng, R.-C. and Zheng, Y.-G. (2023). From discovery to mass production: A perspective on bio-manufacturing exemplified by the development of statins. *Engineering,* **24**, 138-150.

35. Sanchez-Garcia, L., Martin, L., Mangues, R., Ferrer-Miralles, N., Vazquez, E. and Villaverde, A. (2016). Recombinant pharmaceuticals from microbial cells: a 2015 update. *Microbial Cell Factories,* **15**, 33.

36. Bao, Z., Gao, Y., Song, Y., Ding, N., Li, W., Wu, Q., Zhang, X., Zheng, Y., Li, J. and Hu, X. (2024). Construction of an *Escherichia coli* chassis for efficient biosynthesis of human-like N-linked glycoproteins. *Frontiers in Bioengineering and Biotechnology,* **12**, 1370685.

37. Chigira, Y., Oka, T., Okajima, T. and Jigami, Y. (2008). Engineering of a mammalian O-glycosylation pathway in the yeast *Saccharomyces cerevisiae*: production of O-fucosylated epidermal growth factor domains. *Glycobiology,* **18**, 303-314.

38. Geijer, C., Ledesma-Amaro, R. and Tomas-Pejo, E. (2022). Unraveling the potential of non-conventional yeasts in biotechnology. *FEMS Yeast Research,* **22**.

39. Damasceno, L. M., Huang, C., Jr. and Batt, C. A. (2012). Protein secretion in *Pichia pastoris* and advances in protein production. *Applied Microbiology and Biotechnology,* **93**, 31-39.

40. Segal-Kischinevzky, C., Romero-Aguilar, L., Alcaraz, L. D., López-Ortiz, G., Martínez-Castillo, B., Torres-Ramírez, N., Sandoval, G. and González, J. (2022). Yeasts inhabiting extreme environments and their biotechnological applications. *Microorganisms,* **10**, 794.

41. Alsammar, H. and Delneri, D. (2020). An update on the diversity, ecology and biogeography of the Saccharomyces genus. *FEMS Yeast Research,* **20**, foaa013.

42. Pellegrini, L., Cardinali, G. and Martini, A. (1999). Selection of *Saccharomyces cerevisiae* strains able to ferment at supraoptimal temperatures. *Annali di microbiologia ed enzimologia,* **49**, 55-65.

43. Comitini, F., Gobbi, M., Domizio, P., Romani, C., Lencioni, L., Mannazzu, I. and Ciani, M. (2011). Selected non-*Saccharomyces* wine yeasts in controlled multistarter fermentations with *Saccharomyces cerevisiae*. *Food Microbiology,* **28**, 873-882.

44. Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M.,

Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S. and Verstrepen, K. J. (2016). Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell,* **166**, 1397-1410.

45. Steensels, J., Snoek, T., Meersman, E., Picca Nicolino, M., Voordeckers, K. and Verstrepen, K. J. (2014). Improving industrial yeast strains: exploiting natural and artificial diversity. *FEMS Microbiology Reviews,* **38**, 947-995.

46. Benjaphokee, S., Hasegawa, D., Yokota, D., Asvarak, T., Auesukaree, C., Sugiyama, M., Kaneko, Y., Boonchird, C. and Harashima, S. (2012). Highly efficient bioethanol production by a *Saccharomyces cerevisiae* strain with multiple stress tolerance to high temperature, acid and ethanol. *New Biotechnology,* **29**, 379-386.

47. Krishnamoorthy, R. N., Vijila, K. and Kumutha, K. (2010). Intergeneric protoplast fusion of yeast for high ethanol production from cheese industry waste – Whey. *Journal of Yeast and Fungal Research,* **1**, 81-87.

48. Thammasittirong, S. N., Thirasaktana, T., Thammasittirong, A. and Srisodsuk, M. (2013). Improvement of ethanol production by ethanol-tolerant *Saccharomyces cerevisiae* UVNR56. *Springerplus,* **2**, 583.

49. Korona, R. (1999). Unpredictable fitness transitions between haploid and diploid strains of the genetically loaded yeast *Saccharomyces cerevisiae*. *Genetics,* **151**, 77-85.

50. Zörgö, E., Chwialkowska, K., Gjuvsland, A. B., Garré, E., Sunnerhagen, P., Liti, G., Blomberg, A., Omholt, S. W. and Warringer, J. (2013). Ancient evolutionary trade-offs between yeast ploidy states. *PLOS Genetics,* **9**, e1003388.

51. Otto, S. P. and Goldstein, D. B. (1992). Recombination and the evolution of diploidy. *Genetics,* **131**, 745-751.

52. Walsh, G. and Walsh, E. (2022). Biopharmaceutical benchmarks 2022. *Nature Biotechnology,* **40**, 1722-1760.

53. Schwarz, H., Schmittner, M., Duschl, A. and Horejs-Hoeck, J. (2014). Residual endotoxin contaminations in recombinant proteins are sufficient to activate human CD1c+ dendritic cells. *PLoS One,* **9**, e113840.

54. Ma, Y., Lee, C. J. and Park, J. S. (2020). Strategies for optimizing the production of proteins and peptides with multiple disulfide bonds. *Antibiotics (Basel),* **9**, 541.

55. Kulagina, N., Besseau, S., Godon, C., Goldman, G. H., Papon, N. and Courdavault, V. (2021). Yeasts as biopharmaceutical production platforms. *Frontiers in Fungal Biology,* **2**, 733492.

56. Hamilton, S. R. and Gerngross, T. U. (2007). Glycosylation engineering in yeast: the advent of fully humanized yeast. *Current Opinion in Biotechnology,* **18**, 387-392.

57. Wang, G., Huang, M. and Nielsen, J. (2017). Exploring the potential of *Saccharomyces cerevisiae* for biopharmaceutical protein production. *Current Opinion in Biotechnology,* **48**, 77-84.

58. Herskowitz, I. (1988). Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiological reviews,* **52**, 536-553.

59. Lee, C. S. and Haber, J. E. (2015). Mating-type gene switching in *Saccharomyces cerevisiae*. *Microbiology Spectrum,* **3**, Mdna3-0013-2014.

60. Katz Ezov, T., Chang, S. L., Frenkel, Z., Segrè, A. V., Bahalul, M., Murray, A. W., Leu, J. Y., Korol, A. and Kashi, Y. (2010). Heterothallism in *Saccharomyces cerevisiae* isolates from nature: effect of *HO* locus on the mode of reproduction. *Molecular ecology,* **19**, 121-131.

61. Belda, I., Ruiz, J., Santos, A., Van Wyk, N. and Pretorius, I. S. (2019). *Saccharomyces cerevisiae*. *Trends in Genetics,* **35**, 956-957.

62. Peter, J., De Chiara, M., Friedrich, A., Yue, J. X., Pflieger, D., Bergstrom, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J. M., Istace, B., Lebrigand, K.,

Barbry, P., Engelen, S., Lemainque, A., Wincker, P., Liti, G. and Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature,* **556**, 339-344.

63. Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. A., Davey, R. P., Roberts, I. N., Burt, A., Koufopanou, V., Tsai, I. J., Bergman, C. M., Bensasson, D., O'Kelly, M. J., van Oudenaarden, A., Barton, D. B., Bailes, E., Nguyen, A. N., Jones, M., Quail, M. A., Goodhead, I., Sims, S., Smith, F., Blomberg, A., Durbin, R. and Louis, E. J. (2009). Population genomics of domestic and wild yeasts. *Nature,* **458**, 337-341.

64. Bendixsen, D. P., Peris, D. and Stelkens, R. (2021). Patterns of genomic instability in interspecific yeast hybrids with diverse ancestries. *Frontiers in Fungal Biology,* **2**, 742894.

65. Hutzler, M., Morrissey, J. P., Laus, A., Meussdoerffer, F. and Zarnkow, M. (2023). A new hypothesis for the origin of the lager yeast *Saccharomyces pastorianus*. *FEMS Yeast Research,* **23**, foad023.

66. Futcher, A. B. (1988). The 2 micron circle plasmid of *Saccharomyces cerevisiae*. *Yeast,* **4**, 27-40.

67. Chan, K. M., Liu, Y. T., Ma, C. H., Jayaram, M. and Sau, S. (2013). The 2 micron plasmid of *Saccharomyces cerevisiae*: a miniaturized selfish genome with optimized functional competence. *Plasmid,* **70**, 2-17.

68. Futcher, A. B. and Cox, B. S. (1984). Copy number and the stability of 2-micron circle-based artificial plasmids of *Saccharomyces cerevisiae*. *Journal of Bacteriology,* **157**, 283-290.

69. Cubillos, F. A., Parts, L., Salinas, F., Bergstrom, A., Scovacricchi, E., Zia, A., Illingworth, C. J., Mustonen, V., Ibstedt, S., Warringer, J., Louis, E. J., Durbin, R. and Liti, G. (2013). High-resolution mapping of complex traits with a four-parent advanced intercross yeast population. *Genetics,* **195**, 1141-1155.

70. Richardson, S. M., Mitchell, L. A., Stracquadanio, G., Yang, K., Dymond, J. S., DiCarlo, J. E., Lee, D., Huang, C. L. V., Chandrasegaran, S., Cai, Y., Boeke, J. D. and Bader, J. S. (2017). Design of a synthetic yeast genome. *Science,* **355**, 1040-1044.

71. Dymond, J. and Boeke, J. (2012). The *Saccharomyces cerevisiae* SCRaMbLE system and genome minimization. *Bioengineered,* **3**, 170-173.

72. Eigenfeld, M. and Schwaminger, S. P. (2025). Cellular variability as a driver for bioprocess innovation and optimization. *Biotechnology Advances,* **79**, 108528.

73. Anane, E., Sawatzki, A., Neubauer, P. and Cruz-Bournazou, M. N. (2019). Modelling concentration gradients in fed-batch cultivations of *E. coli* – towards the flexible design of scale-down experiments. *Journal of Chemical Technology & Biotechnology,* **94**, 516-526.

74. Bodizs, L., Titica, M., Faria, N., Srinivasan, B., Dochain, D. and Bonvin, D. (2007). Oxygen control for an industrial pilot-scale fed-batch filamentous fungal fermentation. *Journal of Process Control,* **17**, 595-606.

75. Nadal-Rey, G., McClure, D. D., Kavanagh, J. M., Cornelissen, S., Fletcher, D. F. and Gernaey, K. V. (2021). Understanding gradients in industrial bioreactors. *Biotechnology Advances,* **46**, 107660.

76. Fitzgerald, D. M., Hastings, P. J. and Rosenberg, S. M. (2017). Stress-induced mutagenesis: implications in cancer and drug resistance. *Annual Review of Cancer Biology,* **1**, 119-140.

77. Wright, N. R., Jessop-Fabre, M. M., Sanchez, B. J., Wulff, T., Workman, C. T., Ronnest, N. P. and Sonnenschein, N. (2022). Emergence of phenotypically distinct subpopulations is a factor in adaptation of recombinant *Saccharomyces cerevisiae* under glucose-limited conditions. *Applied and Environmental Microbiology,* **88**, e0230721.

78. Worland, A. M., Xu, V. A., Duran, M. F., Gitman, P., Hunter-Cevera, K., Klemm, C., Sun, Y., Sanchis, D. R., Ledesma-Amaro, R., Pomraning, K. R., Tanjore, D., Blenner, M. and Tang, Y. J. (2026). Staying productive under pressure: systems evaluations of β-carotene production in *Yarrowia lipolytica* under continuous fermentation. *Trends in Biotechnology,* **44**, 154-169.

79. Lee, S. W., Rugbjerg, P. and Sommer, M. O. A. (2021). Exploring selective pressure trade-offs for synthetic addiction to extend metabolite productive lifetimes in yeast. *ACS Synthetic Biology,* **10**, 2842-2849.

80. Rugbjerg, P., Sarup-Lytzen, K., Nagy, M. and Sommer, M. O. A. (2018). Synthetic addiction extends the productive life time of engineered *Escherichia coli* populations. *Proceedings of the National Academy of Sciences of the United States of America,* **115**, 2347-2352.

81. Rugbjerg, P., Myling-Petersen, N., Porse, A., Sarup-Lytzen, K. and Sommer, M. O. A. (2018). Diverse genetic error modes constrain large-scale bio-based production. *Nature Communications,* **9**, 787.

82. Galhardo, R. S., J., H. P. and and Rosenberg, S. M. (2007). Mutation as a stress response and the regulation of evolvability. *Critical Reviews in Biochemistry and Molecular Biology,* **42**, 399-435.

83. Rugbjerg, P. and Sommer, M. O. A. (2019). Overcoming genetic heterogeneity in industrial fermentations. *Nature Biotechnology,* **37**, 869-876.

84. Gnugge, R. and Rudolf, F. (2017). *Saccharomyces cerevisiae* shuttle vectors. *Yeast,* **34**, 205-221.

85. Silva, F., Queiroz, J. A. and Domingues, F. C. (2012). Evaluating metabolic stress and plasmid stability in plasmid DNA production by *Escherichia coli*. *Biotechnology Advances,* **30**, 691-708.

86. Hu, K. K. Y., Suri, A., Dumsday, G. and Haritos, V. S. (2024). Cross-feeding promotes heterogeneity within yeast cell populations. *Nature Communications,* **15**, 418.

87. Chen, H. and Jiang, P. (2019). Metabolic engineering of *Escherichia coli* for efficient biosynthesis of fluorescent phycobiliprotein. *Microbial Cell Factories,* **18**, 58.

88. Park, S., You, X. and Imlay, J. A. (2005). Substantial DNA damage from submicromolar intracellular hydrogen peroxide detected in Hpx- mutants of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America,* **102**, 9317-9322.

89. Farr, S. B., D'Ari, R. and Touati, D. (1986). Oxygen-dependent mutagenesis in *Escherichia coli* lacking superoxide dismutase. *Proceedings of the National Academy of Sciences,* **83**, 8268-8272.

90. Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., Dopman, E. B., Dickinson, W. J., Okamoto, K., Kulkarni, S., Hartl, D. L. and Thomas, W. K. (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proceedings of the National Academy of Sciences of the United States of America,* **105**, 9272-9277.

91. Puig, S., Querol, A., Barrio, E. and Perez-Ortin, J. E. (2000). Mitotic recombination and genetic changes in *Saccharomyces cerevisiae* during wine fermentation. *Applied and Environmental Microbiology,* **66**, 2057-2061.

92. Makova, K. D. and Hardison, R. C. (2015). The effects of chromatin organization on variation in mutation rates in the genome. *Nature Reviews Genetics,* **16**, 213-223.

93. Hasenauer, F. C., Barreto, H. C., Lotton, C. and Matic, I. (2025). Genome-wide mapping of spontaneous DNA replication error-hotspots using mismatch repair proteins in rapidly proliferating *Escherichia coli*. *Nucleic Acids Research,* **53**.

94. Suárez, G. A., Renda, B. A., Dasgupta, A., Barrick, J. E. and Kivisaar, M. (2017). Reduced mutation rate and increased transformability of transposon-free *Acinetobacter baylyi* ADP1-ISx. *Applied and Environmental Microbiology,* **83**, e01025-01017.

95. Sleight, S. C., Bartley, B. A., Lieviant, J. A. and Sauro, H. M. (2010). Designing and engineering evolutionary robust genetic circuits. *Journal of Biological Engineering,* **4**, 12.

96. Avery, S. V. (2006). Microbial cell individuality and the underlying sources of heterogeneity. *Nature Reviews Microbiology* **4**, 577-587.

97. Eldar, A. and Elowitz, M. B. (2010). Functional roles for noise in genetic circuits. *Nature,* **467**, 167-173.

98. Booth, I. R. (2002). Stress and the single cell: Intrapopulation diversity is a mechanism to ensure survival upon exposure to stress. *International Journal of Food Microbiology,* **78**, 19-30.

99. Raser, J. M. and O'Shea, E. K. (2005). Noise in gene expression: origins, consequences, and control. *Science,* **309**, 2010-2013.

100. Lin, W. H. and Jacobs-Wagner, C. (2022). Connecting single-cell ATP dynamics to overflow metabolism, cell growth, and the cell cycle in *Escherichia coli*. *Current Biology,* **32**, 3911-3924 e3914.

101. Paige, J. S., Nguyen-Duc, T., Song, W. and Jaffrey, S. R. (2012). Fluorescence imaging of cellular metabolites with RNA. *Science,* **335**, 1194.

102. Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., Emili, A. and Xie, X. S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science,* **329**, 533-538.

103. Grote, J., Krysciak, D. and Streit Wolfgang, R. (2015). Phenotypic heterogeneity, a phenomenon that may explain why quorum sensing does not always result in truly homogenous cell behavior. *Applied and Environmental Microbiology,* **81**, 5280-5289.

104. Kiviet, D. J., Nghe, P., Walker, N., Boulineau, S., Sunderlikova, V. and Tans, S. J. (2014). Stochasticity of metabolism and growth at the single-cell level. *Nature,* **514**, 376-379.

105. Bergmiller, T., Andersson, A. M. C., Tomasek, K., Balleza, E., Kiviet, D. J., Hauschild, R., Tkacik, G. and Guet, C. C. (2017). Biased partitioning of the multidrug efflux pump AcrAB-TolC underlies long-lived phenotypic heterogeneity. *Science,* **356**, 311-315.

106. Sumner, E. R. and Avery, S. V. (2002). Phenotypic heterogeneity: differential stress resistance among individual cells of the yeast *Saccharomyces cerevisiae*. *Microbiology-Sgm,* **148**, 345-351.

107. Engl, C., Jovanovic, G., Brackston, R. D., Kotta-Loizou, I. and Buck, M. (2020). The route to transcription initiation determines the mode of transcriptional bursting in *E. coli*. *Nature Communications,* **11**, 2422.

108. Stewart-Ornstein, J., Weissman, Jonathan S. and El-Samad, H. (2012). Cellular noise regulons underlie fluctuations in *Saccharomyces cerevisiae*. *Molecular Cell,* **45**, 483-493.

109. Soltani, M., Vargas-Garcia, C. A., Antunes, D. and Singh, A. (2016). Intercellular variability in protein levels from stochastic expression and noisy cell cycle processes. *PLOS Computational Biology,* **12**, e1004972.

110. Elowitz, M. B., Levine, A. J., Siggia, E. D. and Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science,* **297**, 1183-1186.

111. Cardinale, S. and Arkin, A. P. (2012). Contextualizing context for synthetic biology – identifying causes of failure of synthetic biological systems. *Biotechnology Journal,* **7**, 856-866.

112. Han, Y. and Zhang, F. (2020). Heterogeneity coordinates bacterial multi-gene expression in single cells. *PLOS Computational Biology,* **16**, e1007643.

113. Yaginuma, H., Kawai, S., Tabata, K. V., Tomiyama, K., Kakizuka, A., Komatsuzaki, T., Noji, H. and Imamura, H. (2014). Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. *Scientific Reports,* **4**, 6522.

114. Beaulaurier, J., Schadt, E. E. and Fang, G. (2019). Deciphering bacterial epigenomes using modern sequencing technologies. *Nature Reviews Genetics,* **20**, 157-172.

115. Jennings, M. P., Hood, D. W., Peak, I. R., Virji, M. and Moxon, E. R. (1995). Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-N-neotetraose terminal lipopolysaccharide structure in *Neisseria meningitidis*. *Molecular Microbiology,* **18**, 729-740.

116. van der Ende, A., Hopman, C. T., Zaat, S., Essink, B. B., Berkhout, B. and Dankert, J. (1995). Variable expression of class 1 outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the -10 and -35 regions of the promoter. *Journal of Bacteriology,* **177**, 2475-2480.

117. Carter, B. and Zhao, K. (2021). The epigenetic basis of cellular heterogeneity. *Nature Reviews Genetics,* **22**, 235-250.

118. O'Kane, C. J. and Hyland, E. M. (2019). Yeast epigenetics: the inheritance of histone modification states. *Bioscience Reports,* **39**.

119. Cusanovich, D. A., Daza, R., Adey, A., Pliner, H. A., Christiansen, L., Gunderson, K. L., Steemers, F. J., Trapnell, C. and Shendure, J. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science,* **348**, 910-914.

120. Halme, A., Bumgarner, S., Styles, C. and Fink, G. R. (2004). Genetic and epigenetic regulation of the *FLO* gene family generates cell-surface variation in yeast. *Cell,* **116**, 405-415.

121. Carrier, F. (2013). Chromatin modulation by histone deacetylase inhibitors: Impact on cellular sensitivity to ionizing radiation. *Molecular and Cellular Pharmacology,* **5**, 51-59.

122. Delvigne, F. and Goffin, P. (2014). Microbial heterogeneity affects bioprocess robustness: dynamic single-cell analysis contributes to understanding of microbial populations. *Biotechnology Journal,* **9**, 61-72.

123. Pinheiro, S., Pandey, S. and Pelet, S. (2022). Cellular heterogeneity: yeast-side story. *Fungal Biology Reviews,* **39**, 34-45.

124. Hao, H., Yao, M., Wang, Y., Zhang, C., Liu, Z., Nielsen, J., Shi, S., Xiao, W. and Yuan, Y. (2024). Extending the G1 phase improves the production of lipophilic compounds in yeast by boosting enzyme expression and increasing cell size. *Proceedings of the National Academy of Sciences of the United States of America,* **121**, e2413486121.

125. Flattery-O'Brien, J. A. and Dawes, I. W. (1998). Hydrogen peroxide causes RAD9-dependent cell cycle arrest in G2 in *Saccharomyces cerevisiae* whereas menadione causes G1 arrest independent of RAD9 function. *Journal of Biological Chemistry,* **273**, 8564-8571.

126. Leroy, C., Mann, C. and Marsolier, M. C. (2001). Silent repair accounts for cell cycle specificity in the signaling of oxidative DNA lesions. *EMBO Journal,* **20**, 2896-2906.

127. Zopf, C. J., Quinn, K., Zeidman, J. and Maheshri, N. (2013). Cell-cycle dependence of transcription dominates noise in gene expression. *PLOS Computational Biology,* **9**, e1003161.

128. Laub, M. T., McAdams, H. H., Feldblyum, T., Fraser, C. M. and Shapiro, L. (2000). Global analysis of the genetic network controlling a bacterial cell cycle. *Science,* **290**, 2144-2148.

129. Weitao, T., Nordstrom, K. and Dasgupta, S. (2000). *Escherichia coli* cell cycle control genes affect chromosome superhelicity. *EMBO Reports,* **1**, 494-499.

130. Bechtloff, D., Grunenfelder, B., Akerlund, T. and Nordstrom, K. (1999). Analysis of protein synthesis rates after initiation of chromosome replication in *Escherichia coli*. *Journal of Bacteriology,* **181**, 6292-6299.

131. Onyeaka, H., Nienow, A. W. and Hewitt, C. J. (2003). Further studies related to the scale-up of high cell density *Escherichia coli* fed-batch fermentations. *Biotechnology and Bioengineering,* **84**, 474-484.

132. Lara, A. R., Galindo, E., Ramírez, O. T. and Palomares, L. A. (2006). Living with heterogeneities in bioreactors. *Molecular Biotechnology,* **34**, 355-381.

133. Kamrad, S., Correia-Melo, C., Szyrwiel, L., Aulakh, S. K., Bahler, J., Demichev, V., Mulleder, M. and Ralser, M. (2023). Metabolic heterogeneity and cross-feeding within isogenic yeast populations captured by DILAC. *Nature Microbiology,* **8**, 441-454.

134. Plocek, V., Váchová, L., Šťovíček, V. and Palková, Z. (2020). Cell distribution within yeast colonies and colony biofilms: How structure develops. *International Journal of Molecular Sciences,* **21**, 3873.

135. Zhong, J.-J. (2010). Recent advances in bioreactor engineering. *Korean Journal of Chemical Engineering,* **27**, 1035-1041.

136. Pörtner, R., Platas Barradas, O., Frahm, B. and Hass, V. C. (2017). Advanced process and control strategies for bioreactors. *Current Developments in Biotechnology and Bioengineering*: Elsevier. 463-493.

137. Everitt, B. and Skrondal, A. (2010). *The Cambridge dictionary of statistics.* 4th edn. Cambridge, UK ; New York: Cambridge University Press.

138. Binder, D., Probst, C., Grunberger, A., Hilgers, F., Loeschcke, A., Jaeger, K. E., Kohlheyer, D. and Drepper, T. (2016). Comparative single-cell analysis of different *E. coli* expression systems during microfluidic cultivation. *PLoS One,* **11**, e0160711.

139. Shabestary, K., Klemm, C., Carling, B., Marshall, J., Savigny, J., Storch, M. and Ledesma-Amaro, R. (2024). Phenotypic heterogeneity follows a growth-viability tradeoff in response to amino acid identity. *Nature Communications,* **15**, 6515.

140. Tonn, M. K., Thomas, P., Barahona, M. and Oyarzun, D. A. (2019). Stochastic modelling reveals mechanisms of metabolic heterogeneity. *Communications Biology,* **2**, 108.

141. Mahdessian, D., Cesnik, A. J., Gnann, C., Danielsson, F., Stenström, L., Arif, M., Zhang, C., Le, T., Johansson, F., Schutten, R., Bäckström, A., Axelsson, U., Thul, P., Cho, N. H., Carja, O., Uhlén, M., Mardinoglu, A., Stadler, C., Lindskog, C., Ayoglu, B., Leonetti, M. D., Pontén, F., Sullivan, D. P. and Lundberg, E. (2021). Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature,* **590**, 649-654.

142. Jacquemart, R., Vandersluis, M., Zhao, M., Sukhija, K., Sidhu, N. and Stout, J. (2016). A single-use strategy to enable manufacturing of affordable biologics. *Computational and Structural Biotechnology Journal,* **14**, 309-318.

143. Frahm, B. (2014). Seed train optimization for cell culture. *Methods in Molecular Biology,* **1104**, 355-367.

144. Henrion, L., Martinez, J. A., Vandenbroucke, V., Delvenne, M., Telek, S., Zicler, A., Grunberger, A. and Delvigne, F. (2023). Fitness cost associated with cell phenotypic switching drives population diversification dynamics and controllability. *Nature Communications,* **14**, 6128.

145. Xiao, Y., Bowen, C. H., Liu, D. and Zhang, F. Z. (2016). Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis. *Nature Chemical Biology,* **12**, 339-346.

146. Kazemi Seresht, A., Cruz, A. L., de Hulster, E., Hebly, M., Palmqvist, E. A., van Gulik, W., Daran, J. M., Pronk, J. and Olsson, L. (2013). Long-term adaptation of *Saccharomyces*

cerevisiae to the burden of recombinant insulin production. *Biotechnology and Bioengineering,* **110**, 2749-2763.

147. Green, S. R. and Moehle, C. M. (1999). Media and culture of yeast. *Current Protocols in Cell Biology,* **4**, 161-162.

148. Sleep, D., Finnis, C., Turner, A. and Evans, L. (2001). Yeast 2 microm plasmid copy number is elevated by a mutation in the nuclear gene *UBC4*. *Yeast,* **18**, 403-421.

149. Gietz, R. D. and Woods, R. A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods in Enzymology,* **350**, 87-96.

150. Malci, K., Watts, E., Roberts, T. M., Auxillos, J. Y., Nowrouzi, B., Boll, H. O., Nascimento, C., Andreou, A., Vegh, P., Donovan, S., Fragkoudis, R., Panke, S., Wallace, E., Elfick, A. and Rios-Solis, L. (2022). Standardization of synthetic biology tools and assembly methods for *Saccharomyces cerevisiae* and emerging yeast species. *ACS Synthetic Biology,* **11**, 2527-2547.

151. Lian, J., HamediRad, M. and Zhao, H. (2018). Advancing metabolic engineering of *Saccharomyces cerevisiae* using the CRISPR/Cas system. *Biotechnology Journal,* **13**, e1700601.

152. Lian, J., Jin, R. and Zhao, H. (2016). Construction of plasmids with tunable copy numbers in *Saccharomyces cerevisiae* and their applications in pathway optimization and multiplex genome integration. *Biotechnology and Bioengineering,* **113**, 2462-2473.

153. Kang, C. W., Lim, H. G., Yang, J., Noh, M. H., Seo, S. W. and Jung, G. Y. (2018). Synthetic auxotrophs for stable and tunable maintenance of plasmid copy number. *Metabolic Engineering,* **48**, 121-128.

154. Klein, T., Niklas, J. and Heinzle, E. (2015). Engineering the supply chain for protein production/secretion in yeasts and mammalian cells. *Journal of Industrial Microbiology and Biotechnology,* **42**, 453-464.

155. Zahrl, R. J., Gasser, B., Mattanovich, D. and Ferrer, P. (2019). Detection and elimination of cellular bottlenecks in protein-producing yeasts. *Methods in Molecular Biology,* **1923**, 75-95.

156. Barton, M. D., Delneri, D., Oliver, S. G., Rattray, M. and Bergman, C. M. (2010). Evolutionary systems biology of amino acid biosynthetic cost in yeast. *PLoS One,* **5**, e11935.

157. Glick, B. R. (1995). Metabolic load and heterologous gene expression. *Biotechnology Advances,* **13**, 247-261.

158. Kastberg, L. L. B., Ard, R., Jensen, M. K. and Workman, C. T. (2022). Burden imposed by heterologous protein production in two major industrial yeast cell factories: identifying sources and mitigation strategies. *Frontiers in Fungal Biology,* **3**, 827704.

159. de Ruijter, J. C., Koskela, E. V., Nandania, J., Frey, A. D. and Velagapudi, V. (2018). Understanding the metabolic burden of recombinant antibody production in *Saccharomyces cerevisiae* using a quantitative metabolomics approach. *Yeast,* **35**, 331-341.

160. Kvitek, D. J., Will, J. L. and Gasch, A. P. (2008). Variations in stress sensitivity and genomic expression in diverse *S. cerevisiae* isolates. *PLOS Genetics,* **4**, e1000223.

161. Almeida, I. C., Pacheco, T. F., Machado, F. and Goncalves, S. B. (2022). Evaluation of different strains of Saccharomyces cerevisiae for ethanol production from high-amylopectin BRS AG rice (Oryza sativa L.). *Scientific Reports,* **12**, 2122.

162. Wong, R. W. K., Foo, M., Lay, J. R. S., Wai, T. L. T., Moore, J., Dutreux, F., Molzahn, C., Nislow, C., Measday, V., Schacherer, J. and Mayor, T. (2025). Mining yeast diversity unveils novel targets for improved heterologous laccase production in Saccharomyces cerevisiae. *Microbial Cell Factories,* **24**, 60.

163. Cadiere, A., Ortiz-Julien, A., Camarasa, C. and Dequin, S. (2011). Evolutionary engineered *Saccharomyces cerevisiae* wine yeast strains with increased in vivo flux through the pentose phosphate pathway. *Metabolic Engineering,* **13**, 263-271.

164. Zhou, K., Qiao, K. J., Edgar, S. and Stephanopoulos, G. (2015). Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nature Biotechnology,* **33**, 377-383.

165. Xu, M., Vallieres, C., Finnis, C., Winzer, K. and Avery, S. V. (2024). Exploiting phenotypic heterogeneity to improve production of glutathione by yeast. *Microbial Cell Factories,* **23**, 267.

166. Davison, S. A., den Haan, R. and van Zyl, W. H. (2020). Exploiting strain diversity and rational engineering strategies to enhance recombinant cellulase secretion by *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology,* **104**, 5163-5184.

167. Sangsoda, S., Cherest, H. and Surdin-Kerjan, Y. (1985). The expression of the MET25 gene of *Saccharomyces cerevisiae* is regulated transcriptionally. *Molecular Genetics and Genomics,* **200**, 407-414.

168. Todd, R. T., Braverman, A. L. and Selmecki, A. (2018). Flow cytometry analysis of fungal ploidy. *Current Protocols in Microbiology,* **50**, e58.

169. Borneman, A. R., Desany, B. A., Riches, D., Affourtit, J. P., Forgan, A. H., Pretorius, I. S., Egholm, M. and Chambers, P. J. (2011). Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLOS Genetics,* **7**, e1001287.

170. Liu, Z., Hou, J., Martinez, J. L., Petranovic, D. and Nielsen, J. (2013). Correlation of cell growth and heterologous protein production by *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology,* **97**, 8955-8962.

171. Li, F., Chen, Y., Qi, Q., Wang, Y., Yuan, L., Huang, M., Elsemman, I. E., Feizi, A., Kerkhoven, E. J. and Nielsen, J. (2022). Improving recombinant protein production by yeast through genome-scale modeling using proteome constraints. *Nature Communications,* **13**, 2969.

172. Nishant, K. T., Wei, W., Mancera, E., Argueso, J. L., Schlattl, A., Delhomme, N., Ma, X., Bustamante, C. D., Korbel, J. O., Gu, Z., Steinmetz, L. M. and Alani, E. (2010). The baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis. *PLOS Genetics,* **6**, e1001109.

173. Mead, D. J., Gardner, D. C. J. and Oliver, S. G. (1986). Enhanced stability of a 2μ-based recombinant plasmid in diploid yeast. *Biotechnology Letters,* **8**, 391-396.

174. Dawes, I. W. and Hardie, I. D. (1974). Selective killing of vegetative cells in sporulated yeast cultures by exposure to diethyl ether. *Molecular Genetics and Genomics,* **131**, 281-289.

175. Hou, J., Tyo, K., Liu, Z., Petranovic, D. and Nielsen, J. (2012). Engineering of vesicle trafficking improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Metabolic Engineering,* **14**, 120-127.

176. Huang, M., Bao, J., Hallstrom, B. M., Petranovic, D. and Nielsen, J. (2017). Efficient protein production by yeast requires global tuning of metabolism. *Nature Communications,* **8**, 1131.

177. Chen, A., Si, Q., Xu, Q., Pan, C., Qu, T. and Chen, J. (2025). Evaluation of stress tolerance and fermentation performance in commercial yeast strains for industrial applications. *Foods,* **14**, 142.

178. Duan, X., Sarangi, P., Liu, X., Rangi, G. K., Zhao, X. and Ye, H. (2009). Structural and functional insights into the roles of the Mms21 subunit of the Smc5/6 complex. *Molecular Cell,* **35**, 657-668.

179. Bermúdez-López, M., Pociño-Merino, I., Sánchez, H., Bueno, A., Guasch, C., Almedawar, S., Bru-Virgili, S., Garí, E., Wyman, C., Reverter, D., Colomina, N. and Torres-Rosell, J. (2015). ATPase-dependent control of the *MMS21* SUMO ligase during DNA repair. *PLOS Biology,* **13**, e1002089.

180. Hays, M., Young, J. M., Levan, P. F. and Malik, H. S. (2020). A natural variant of the essential host gene *MMS21* restricts the parasitic 2-micron plasmid in *Saccharomyces cerevisiae*. *Elife,* **9**, e62337.

181. Jayakrishnan, A., Wan Rosli, W. R., Tahir, A. R. M., Razak, F. S. A., Kee, P. E., Ng, H. S., Chew, Y.-L., Lee, S.-K., Ramasamy, M., Tan, C. S. and Liew, K. B. (2024). Evolving paradigms of pecombinant protein production in pharmaceutical industry: a rigorous review. *Sci,* **6**, 9.

182. Carroll, A. C. and Wong, A. (2018). Plasmid persistence: costs, benefits, and the plasmid paradox. *Canadian Journal of Microbiology,* **64**, 293-304.

183. Duveau, F., Yuan, D. C., Metzger, B. P. H., Hodgins-Davis, A. and Wittkopp, P. J. (2017). Effects of mutation and selection on plasticity of a promoter activity in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America,* **114**, e11218-e11227.

184. Vashist, S. and Ng, D. T. (2004). Misfolded proteins are sorted by a sequential checkpoint mechanism of ER quality control. *Journal of Cell Biology,* **165**, 41-52.

185. Isken, O. and Maquat, L. E. (2008). The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nature Reviews Genetics,* **9**, 699-712.

186. Peltz, S. W., Brown, A. H. and Jacobson, A. (1993). mRNA destabilization triggered by premature translational termination depends on at least three cis-acting sequence elements and one trans-acting factor. *Genes & development,* **7**, 1737-1754.

187. Gresham, D. and Hong, J. (2015). The functional basis of adaptive evolution in chemostats. *FEMS Microbiology Reviews,* **39**, 2-16.

188. Jahn, M., Vorpahl, C., Hübschmann, T., Harms, H. and Müller, S. (2016). Copy number variability of expression plasmids determined by cell sorting and droplet digital PCR. *Microbial Cell Factories,* **15**, 211.

189. Hohnholz, R., Pohlmann, K. J. and Achstetter, T. (2017). A set of isomeric episomal plasmids for systematic examination of mitotic stability in *Saccharomyces cerevisiae*. *Yeast,* **34**, 267-275.

190. Clark-Walker, G. D. and Miklos, G. L. (1974). Localization and quantification of circular DNA in yeast. *European Journal of Biochemistry,* **41**, 359-365.

191. Gerbaud, C. and Guerineau, M. (1980). 2 micron plasmid copy number in different yeast strains and repartition of endogenous and 2 micron chimeric plasmids in transformed strains. *Current Genetics,* **1**, 219-228.

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

193. Ghosh, S. K., Hajra, S. and Jayaram, M. (2007). Faithful segregation of the multicopy yeast plasmid through cohesin-mediated recognition of sisters. *Proceedings of the National Academy of Sciences of the United States of America,* **104**, 13034-13039.

194. Kumar, D. and Ghosh, S. K. (2024). Chromosome hitchhiking: a potential strategy adopted by the selfish yeast plasmids to ensure symmetric inheritance during cell division. *Biochemical Society Transactions,* **52**, 2359-2372.

195. Mei, H., Arbeithuber, B., Cremona, M. A., DeGiorgio, M. and Nekrutenko, A. (2019). A high-resolution view of adaptive event dynamics in a plasmid. *Genome Biology and Evolution,* **11**, 3022-3034.

196. Wright, R. C. T., Wood, A. J., Bottery, M. J., Muddiman, K. J., Paterson, S., Harrison, E., Brockhurst, M. A. and Hall, J. P. J. (2024). A chromosomal mutation is superior to a plasmid-encoded mutation for plasmid fitness cost compensation. *PLOS Biology,* **22**, e3002926.

197. Singh, M. V. and Weil, P. A. (2002). A method for plasmid purification directly from yeast. *Analytical Biochemistry,* **307**, 13-17.

198. Sherman, F. (2002). Getting started with yeast. *Methods in Enzymology,* **350**, 3-41.

199. Stovicek, V., Vachova, L., Begany, M., Wilkinson, D. and Palkova, Z. (2014). Global changes in gene expression associated with phenotypic switching of wild yeast. *BMC Genomics,* **15**, 136.

200. Ilhan, J., Kupczok, A., Woehle, C., Wein, T., Hülter, N. F., Rosenstiel, P., Landan, G., Mizrahi, I. and Dagan, T. (2019). Segregational drift and the interplay between plasmid copy number and evolvability. *Molecular Biology and Evolution,* **36**, 472-486.

201. Foster, P. L. (2007). Stress-induced mutagenesis in bacteria. *Critical Reviews in Biochemistry and Molecular Biology,* **42**, 373-397.

202. MacLean, R. C., Torres-Barceló, C. and Moxon, R. (2013). Evaluating evolutionary models of stress-induced mutagenesis in bacteria. *Nature Reviews Genetics,* **14**, 221-227.

203. Peng, B., Williams, T. C., Henry, M., Nielsen, L. K. and Vickers, C. E. (2015). Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. *Microbial Cell Factories,* **14**, 91.

204. Friehs, K. (2004). Plasmid Copy Number and Plasmid Stability. in Scheper, T. (ed.) *New Trends and Developments in Biochemical Engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg. 47-82.

205. Liu, Y. T., Sau, S., Ma, C. H., Kachroo, A. H., Rowley, P. A., Chang, K. M., Fan, H. F. and Jayaram, M. (2014). The partitioning and copy number control systems of the selfish yeast plasmid: an optimized molecular design for stable persistence in host cells. *Microbiology Spectrum,* **2**, 00032013.

206. Senecoff, J. F., Bruckner, R. C. and Cox, M. M. (1985). The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. *Proceedings of the National Academy of Sciences of the United States of America,* **82**, 7270-7274.

207. Prajapati, H. K., Rizvi, S. M. A., Rathore, I. and Ghosh, S. K. (2017). Microtubule-associated proteins, Bik1 and Bim1, are required for faithful partitioning of the endogenous 2 micron plasmids in budding yeast. *Molecular Microbiology,* **103**, 1046-1064.

208. Sau, S., Conrad, M. N., Lee, C.-Y., Kaback, D. B., Dresser, M. E. and Jayaram, M. (2014). A selfish DNA element engages a meiosis-specific motor and telomeres for germ-line propagation. *Journal of Cell Biology,* **205**, 643-661.

209. Zakian, V. A., Brewer, B. J. and Fangman, W. L. (1979). Replication of each copy of the yeast 2 micron DNA plasmid occurs during the S phase. *Cell,* **17**, 923-934.

210. Gresham, D., Desai, M. M., Tucker, C. M., Jenq, H. T., Pai, D. A., Ward, A., DeSevo, C. G., Botstein, D. and Dunham, M. J. (2008). The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLOS Genetics,* **4**, e1000303.

211. Bodi, Z., Farkas, Z., Nevozhay, D., Kalapis, D., Lazar, V., Csorgo, B., Nyerges, A., Szamecz, B., Fekete, G., Papp, B., Araujo, H., Oliveira, J. L., Moura, G., Santos, M. A. S., Szekely, T., Jr.,

Balazsi, G. and Pal, C. (2017). Phenotypic heterogeneity promotes adaptive evolution. *PLOS Biology,* **15**, e2000644.

212. Cao, Y., Wu, Y., Lv, X., Li, J., Liu, L., Du, G., Chen, J. and Liu, Y. (2025). Reduced genetic heterogeneity for stable bioproduction by harnessing the bias and mechanism of mutation. *Microbial Biotechnology,* **18**, e70162.

213. Smith, M. C., Sumner, E. R. and Avery, S. V. (2007). Glutathione and Gts1p drive beneficial variability in the cadmium resistances of individual yeast cells. *Molecular Microbiology,* **66**, 699-712.

214. Holland, S. L., Reader, T., Dyer, P. S. and Avery, S. V. (2014). Phenotypic heterogeneity is a selected trait in natural yeast populations subject to environmental stress. *Environmental Microbiology,* **16**, 1729-1740.

215. Levy, S. F., Ziv, N. and Siegal, M. L. (2012). Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLOS Biology,* **10**, e1001325.

216. Lukačišin, M., Espinosa-Cantú, A. and Bollenbach, T. (2022). Intron-mediated induction of phenotypic heterogeneity. *Nature,* **605**, 113-118.

217. Lang, G. I., Rice, D. P., Hickman, M. J., Sodergren, E., Weinstock, G. M., Botstein, D. and Desai, M. M. (2013). Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature,* **500**, 571-574.

218. Levy, S. F., Blundell, J. R., Venkataram, S., Petrov, D. A., Fisher, D. S. and Sherlock, G. (2015). Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature,* **519**, 181-186.

219. Kurland, C. and Gallant, J. (1996). Errors of heterologous protein expression. *Current Opinion in Biotechnology,* **7**, 489-493.

220. Snoeck, S., Guidi, C. and De Mey, M. (2024). "Metabolic burden" explained: stress symptoms and its related responses induced by overexpression of heterologous proteins in *Escherichia coli*. *Microbial Cell Factories,* **23**, 96.

221. Quan, N., Eguchi, Y. and Geiler-Samerotte, K. (2023). Intra-FCY1: a novel system to identify mutations that cause protein misfolding. *Frontiers in Genetics,* **14**, 1198203.

222. Kuras, L., Barbey, R. and Thomas, D. (1997). Assembly of a bZIP-bHLH transcription activation complex: formation of the yeast Cbf1-Met4-Met28 complex is regulated through Met28 stimulation of Cbf1 DNA binding. *EMBO Journal,* **16**, 2441-2451.

223. Thomas, D., Jacquemin, I. and Surdin-Kerjan, Y. (1992). MET4, a leucine zipper protein, and centromere-binding factor 1 are both required for transcriptional activation of sulfur metabolism in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology,* **12**, 1719-1727.

224. Mordaka, P. M. and Heap, J. T. (2018). Stringency of synthetic promoter sequences in *Clostridium* revealed and circumvented by tuning promoter library mutation rates. *ACS Synthetic Biology,* **7**, 672-681.

225. He, S., Zhang, Z. and Lu, W. (2023). Natural promoters and promoter engineering strategies for metabolic regulation in *Saccharomyces cerevisiae*. *Journal of Industrial Microbiology and Biotechnology,* **50**, kuac029.

226. Einarsson, H., Salvatore, M., Vaagensø, C., Alcaraz, N., Bornholdt, J., Rennie, S. and Andersson, R. (2022). Promoter sequence and architecture determine expression variability and confer robustness to genetic variants. *Elife,* **11**, e80943.

227. Davis, J. H., Rubin, A. J. and Sauer, R. T. (2011). Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Research,* **39**, 1131-1141.

228. Senecoff, J. F. and Cox, M. M. (1986). Directionality in FLP protein-promoted site-specific recombination is mediated by DNA-DNA pairing. *Journal of Biological Chemistry,* **261**, 7380-7386.

229. Turan, S., Kuehle, J., Schambach, A., Baum, C. and Bode, J. (2010). Multiplexing RMCE: versatile extensions of the Flp-recombinase-mediated cassette-exchange technology. *Journal of Molecular Biology,* **402**, 52-69.

230. Schlake, T. and Bode, J. (1994). Use of mutated FLP recognition target (*FRT*) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry,* **33**, 12746-12751.

231. Voziyanov, Y., Konieczka, J. H., Francis Stewart, A. and Jayaram, M. (2003). Stepwise manipulation of DNA specificity in Flp recombinase: Progressively adapting Flp to individual and combinatorial mutations in its target site. *Journal of Molecular Biology,* **326**, 65-76.

232. Broach, J. R., Atkins, J. F., McGill, C. and Chow, L. (1979). Identification and mapping of the transcriptional and translational products of the yeast plasmid, 2mu circle. *Cell,* **16**, 827-839.

233. Karim, A. S., Curran, K. A. and Alper, H. S. (2013). Characterization of plasmid burden and copy number in *Saccharomyces cerevisiae* for optimization of metabolic engineering applications. *FEMS Yeast Research,* **13**, 107-116.

234. Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H. and Kompala, D. S. (1990). Plasmid-encoded protein: The principal factor in the “metabolic burden” associated with recombinant bacteria. *Biotechnology and Bioengineering,* **35**, 668-681.

235. Newman, J. R., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L. and Weissman, J. S. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature,* **441**, 840-846.

236. Mattanovich, D., Gasser, B., Hohenblum, H. and Sauer, M. (2004). Stress in recombinant protein producing yeasts. *Journal of Biotechnology,* **113**, 121-135.

237. Ogawa, N. and Mori, K. (2004). Autoregulation of the HAC1 gene is required for sustained activation of the yeast unfolded protein response. *Genes to Cells,* **9**, 95-104.

238. Hernández-Elvira, M., Torres-Quiroz, F., Escamilla-Ayala, A., Domínguez-Martin, E., Escalante, R., Kawasaki, L., Ongay-Larios, L. and Coria, R. (2018). The unfolded protein response pathway in the yeast *Kluyveromyces lactis*. A comparative view among yeast species. *Cells,* **7**, 106.

239. Nguyen Phuong Thi, M., Ishiwata-Kimata, Y. and Kimata, Y. (2022). Fast-growing *Saccharomyces cerevisiae* cells with a constitutive unfolded protein response and their potential for lipidic molecule production. *Applied and Environmental Microbiology,* **88**, e01083-01022.

240. Martinez, J. L., Meza, E., Petranovic, D. and Nielsen, J. (2016). The impact of respiration and oxidative stress response on recombinant alpha-amylase production by *Saccharomyces cerevisiae*. *Metabolic Engineering Communications,* **3**, 205-210.

241. Feng, T., Yu, H. and Ye, L. (2025). Mechanisms and strategies for engineering oxidative stress resistance in *Saccharomyces cerevisiae*. *Chem & Bio Engineering,* **10**, cbe.5c00021.

242. Jamieson, D. J. (1998). Oxidative stress responses of the yeast *Saccharomyces cerevisiae*. *Yeast,* **14**, 1511-1527.

243. Delic, M., Graf, A. B., Koellensperger, G., Haberhauer-Troyer, C., Hann, S., Mattanovich, D. and Gasser, B. (2014). Overexpression of the transcription factor Yap1 modifies intracellular redox conditions and enhances recombinant protein secretion. *Microbial Cell,* **1**, 376-386.

244. Li, W. F., Ji, J., Wang, G., Wang, H. Y., Niu, B. L. and Josine, T. L. (2011). Oxidative stress-resistance assay for screening yeast strains overproducing heterologous proteins. *Russian Journal of Genetics,* **47**, 1039-1046.

245. Peng, K., Kroukamp, H., Pretorius, I. S. and Paulsen, I. T. (2021). Yeast Synthetic Minimal Biosensors for Evaluating Protein Production. *ACS Synthetic Biology,* **10**, 1640-1650.

246. Mateus, C. and Avery, S. V. (2000). Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast,* **16**, 1313-1323.

247. James, J., Fiji, N., Roy, D., Andrew Mg, D., Shihabudeen, M. S., Chattopadhyay, D. and Thirumurugan, K. (2015). A rapid method to assess reactive oxygen species in yeast using H2DCF-DA. *Analytical Methods,* **7**, 8572-8575.

248. Sumner, E. R. and Avery, S. V. (2002). Phenotypic heterogeneity: differential stress resistance among individual cells of the yeast *Saccharomyces cerevisiae*. *Microbiology,* **148**, 345-351.

249. Bishop, A. L., Rab, F. A., Sumner, E. R. and Avery, S. V. (2007). Phenotypic heterogeneity can enhance rare-cell survival in 'stress-sensitive' yeast populations. *Molecular Microbiology,* **63**, 507-520.

250. Nadal-Ribelles, M., Lieb, G., Solé, C., Matas, Y., Szachnowski, U., Andjus, S., Quintana, M., Romo, M., Herrero, A. G., Morillon, A., Pelet, S., de Nadal, E. and Posas, F. (2025). Transcriptional heterogeneity shapes stress-adaptive responses in yeast. *Nature Communications,* **16**, 2631.

251. Knafo, M., Rezenman, S., Nevo, R., Tsigalnitski, I., Reich, Z. and Kapon, R. (2025). Variability in protein expression and fitness under stress: distinct modes of early adaptation of populations. *PRX Life,* **3**, 013006.

252. Battye, F. L., Light, A. and Tarlinton, D. M. (2000). Single cell sorting and cloning. *Journal of Immunological Methods,* **243**, 25-32.

253. Young, C. L. and Robinson, A. S. (2014). Protein folding and secretion: mechanistic insights advancing recombinant protein production in *S. cerevisiae*. *Current Opinion in Biotechnology,* **30**, 168-177.

254. Tyo, K. E., Liu, Z., Petranovic, D. and Nielsen, J. (2012). Imbalance of heterologous protein folding and disulfide bond formation rates yields runaway oxidative stress. *BMC Biology,* **10**, 16.

255. Čáp, M. and Palková, Z. (2024). The characteristics of differentiated yeast subpopulations depend on their lifestyle and available nutrients. *Scientific Reports,* **14**, 3681.

256. Pelet, S. and Peter, M. (2011). Dynamic processes at stress promoters regulate the bimodal expression of *HOG* response genes. *Communicative & Integrative Biology,* **4**, 699-702.

257. Zacharioudakis, I. and Tzamarias, D. (2017). Bimodal expression of yeast *GAL* genes is controlled by a long non-coding RNA and a bifunctional galactokinase. *Biochemical and Biophysical Research Communications,* **486**, 63-69.

258. Palme, J., Wang, J. and Springer, M. (2021). Variation in the modality of a yeast signaling pathway is mediated by a single regulator. *Elife,* **10**, e69974.

259. Bheda, P., Aguilar-Gómez, D., Kukhtevich, I., Becker, J., Charvin, G., Kirmizis, A. and Schneider, R. (2020). Microfluidics for single-cell lineage tracking over time to characterize transmission of phenotypes in Saccharomyces cerevisiae. *STAR Protocols,* **1**, 100228.

260. Griesemer, M., Young, C., Robinson, A. S. and Petzold, L. (2014). BiP clustering facilitates protein folding in the endoplasmic reticulum. *PLOS Computational Biology,* **10**, e1003675.

261. Guerra-Moreno, A., Ang, J., Welsch, H., Jochem, M. and Hanna, J. (2019). Regulation of the unfolded protein response in yeast by oxidative stress. *FEBS Letters,* **593**, 1080-1088.

262. Su, Y., Xu, C., Shea, J., DeStephanis, D. and Su, Z. (2023). Transcriptomic changes in single yeast cells under various stress conditions. *BMC Genomics,* **24**, 88.

263. Zechmann, B., Liou, L. C., Koffler, B. E., Horvat, L., Tomasic, A., Fulgosi, H. and Zhang, Z. (2011). Subcellular distribution of glutathione and its dynamic changes under oxidative stress in the yeast *Saccharomyces cerevisiae*. *FEMS Yeast Research,* **11**, 631-642.

264. Want, A., Thomas, O. R., Kara, B., Liddell, J. and Hewitt, C. J. (2009). Studies related to antibody fragment (Fab) production in *Escherichia coli* W3110 fed-batch fermentation processes using multiparameter flow cytometry. *Cytometry Part A,* **75**, 148-154.

265. Hohenblum, H., Borth, N. and Mattanovich, D. (2003). Assessing viability and cell-associated product of recombinant protein producing *Pichia pastoris* with flow cytometry. *Journal of Biotechnology,* **102**, 281-290.

266. Compagno, C., Porro, D., Radice, S., Martegani, E. and Ranzi, B. M. (1996). Selection of yeast cells with a higher plasmid copy number in a *Saccharomyces cerevisiae* autoselection system. *Yeast,* **12**, 199-205.

267. D'Ambrosio, V., Dore, E., Di Blasi, R., van den Broek, M., Sudarsan, S., Horst, J. T., Ambri, F., Sommer, M. O. A., Rugbjerg, P., Keasling, J. D., Mans, R. and Jensen, M. K. (2020). Regulatory control circuits for stabilizing long-term anabolic product formation in yeast. *Metabolic Engineering,* **61**, 369-380.

268. Satyanarayana, T. and Kunze, G. (eds.) (2009). *Yeast biotechnology: diversity and applications*. Springer Dordrecht.

269. Udeh, K. O. and Achremowicz, B. (1997). High-glutathione containing yeast *Saccharomyces cerevisiae*: optimization of production. *Acta Microbiologica Polonica* **46**, 105-114.

270. Kresnowati, M., Ikhsan, N. A., Nursa'adah, R. S., Santoso, N. N. and Susanto, Y. W. (2019). Evaluation of glutathione production method using *Saccharomyces cerevisiae*. *IOP Conference Series: Materials Science and Engineering,* **543**, 012004.

271. Kiriyama, K., Hara, K. Y. and Kondo, A. (2012). Extracellular glutathione fermentation using engineered *Saccharomyces cerevisiae* expressing a novel glutathione exporter. *Applied Microbiology and Biotechnology,* **96**, 1021-1027.

272. Kobayashi, J., Sasaki, D., Hara, K. Y., Hasunuma, T. and Kondo, A. (2022). Metabolic engineering of the L-serine biosynthetic pathway improves glutathione production in *Saccharomyces cerevisiae*. *Microbial Cell Factories,* **21**, 153.

273. Schmacht, M., Lorenz, E., Stahl, U. and Senz, M. (2017). Medium optimization based on yeast's elemental composition for glutathione production in *Saccharomyces cerevisiae*. *Journal of Bioscience and Bioengineering,* **123**, 555-561.

274. Zhang, T., Wen, S. H. and Tan, T. W. (2007). Optimization of the medium for glutathione production in *Saccharomyces cerevisiae*. *Process Biochemistry,* **42**, 454-458.

275. Alfafara, C., Miura, K., Shimizu, H., Shioya, S. and Suga, K. (1992). Cysteine addition strategy for maximum glutathione production in fed-batch culture of *Saccharomyces cerevisiae*. *Applied Microbiology and Biotechnology,* **37**, 141-146.

276. Wheeler, G. L., Quinn, K. A., Perrone, G., Dawes, I. W. and Grant, C. M. (2002). Glutathione regulates the expression of gamma-glutamylcysteine synthetase via the Met4 transcription factor. *Molecular Microbiology,* **46**, 545-556.

277. Wheeler, G. L., Trotter, E. W., Dawes, I. W. and Grant, C. M. (2003). Coupling of the transcriptional regulation of glutathione biosynthesis to the availability of glutathione and methionine via the Met4 and Yap1 transcription factors. *Journal of Biological Chemistry,* **278**, 49920-49928.

278. Rahman, I., Kode, A. and Biswas, S. K. (2006). Assay for quantitative determination of glutathione and glutathione disulfide levels using enzymatic recycling method. *Nature Protocols,* **1**, 3159-3165.

279. Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-ΔΔ C(t)) method. *Methods,* **25**, 402-408.

280. Siewers, V. (2022). An overview on selection marker genes for transformation of *Saccharomyces cerevisiae*. *Methods in Molecular Biology,* **2513**, 1-13.

281. Alani, E., Cao, L. and Kleckner, N. (1987). A method for gene disruption that allows repeated use of Ura3 Selection in the construction of multiply disrupted yeast strains. *Genetics,* **116**, 541-545.

282. Wach, A., Brachat, A., Alberti-Segui, C., Rebischung, C. and Philippsen, P. (1997). Heterologous *HIS3* marker and GFP reporter modules for PCR-targeting in *Saccharomyces cerevisiae*. *Yeast,* **13**, 1065-1075.

283. Sampaio, N. M. V., Ajith, V. P., Watson, R. A., Heasley, L. R., Chakraborty, P., Rodrigues-Prause, A., Malc, E. P., Mieczkowski, P. A., Nishant, K. T. and Argueso, J. L. (2020). Characterization of systemic genomic instability in budding yeast. *Proceedings of the National Academy of Sciences of the United States of America,* **117**, 28221-28231.

284. Regenberg, B. and Hansen, J. (2000). GAP1, a novel selection and counter-selection marker for multiple gene disruptions in *Saccharomyces cerevisiae*. *Yeast,* **16**, 1111-1119.

285. Grenson, M., Hou, C. and Crabeel, M. (1970). Multiplicity of the amino acid permeases in *Saccharomyces cerevisiae*. IV. Evidence for a general amino acid permease. *Journal of Bacteriology,* **103**, 770-777.

286. Jauniaux, J. C. and Grenson, M. (1990). GAP1, the general amino acid permease gene of *Saccharomyces cerevisiae*. Nucleotide sequence, protein similarity with the other bakers yeast amino acid permeases, and nitrogen catabolite repression. *European Journal of Biochemistry,* **190**, 39-44.

287. Lipson, D., Raz, T., Kieu, A., Jones, D. R., Giladi, E., Thayer, E., Thompson, J. F., Letovsky, S., Milos, P. and Causey, M. (2009). Quantification of the yeast transcriptome by single-molecule sequencing. *Nature Biotechnology,* **27**, 652-658.

288. Yim, E., O'Connell, K. E., St Charles, J. and Petes, T. D. (2014). High-resolution mapping of two types of spontaneous mitotic gene conversion events in *Saccharomyces cerevisiae*. *Genetics,* **198**, 181-192.

289. Suzuki, T., Yokoyama, A., Tsuji, T., Ikeshima, E., Nakashima, K., Ikushima, S., Kobayashi, C. and Yoshida, S. (2011). Identification and characterization of genes involved in glutathione production in yeast. *Journal of Bioscience and Bioengineering,* **112**, 107-113.

290. Hu, X., Shen, X., Zhu, S., Zeng, H. and Shuai, Y. (2023). Optimization of glutathione production in *Saccharomyces cerevisiae* HBSD-W08 using Plackett-Burman and central composite rotatable designs. *BMC Microbiology,* **23**, 11.

291. Wang, X. and Marchisio, M. A. (2021). Synthetic polycistronic sequences in eukaryotes. *Synthetic and Systems Biotechnology,* **6**, 254-261.

292. Souza-Moreira, T. M., Navarrete, C., Chen, X., Zanelli, C. F., Valentini, S. R., Furlan, M., Nielsen, J. and Krivoruchko, A. (2018). Screening of 2A peptides for polycistronic gene expression in yeast. *FEMS Yeast Research,* **18**, foy036.

293. Wu, J., Elliston, A., Le Gall, G., Colquhoun, I. J., Collins, S. R. A., Dicks, J., Roberts, I. N. and Waldron, K. W. (2017). Yeast diversity in relation to the production of fuels and chemicals. *Scientific Reports,* **7**, 14259.

294. Deparis, Q., Claes, A., Foulquié-Moreno, M. R. and Thevelein, J. M. (2017). Engineering tolerance to industrially relevant stress factors in yeast cell factories. *FEMS Yeast Research,* **17**.

295. Oman, M., Alam, A. and Ness, R. W. (2022). How sequence context-dependent mutability drives mutation rate variation in the genome. *Genome Biology and Evolution,* **14**.

296. Wang, C., Poborsky, M., Crocoll, C., Nødvig, C. S., Mortensen, U. H. and Halkier, B. A. (2022). Comparison of genome and plasmid-based engineering of multigene benzylglucosinolate pathway in *Saccharomyces cerevisiae*. *Applied and Environmental Microbiology,* **88**, e0097822.

297. Bonny, A. R., Kochanowski, K., Diether, M. and El-Samad, H. (2021). Stress-induced growth rate reduction restricts metabolic resource utilization to modulate osmo-adaptation time. *Cell Reports,* **34**, 108854.

298. Qi, Y., Xu, N., Li, Z., Wang, J., Meng, X., Gao, C., Chen, J., Chen, W., Chen, X. and Liu, L. (2022). Mediator engineering of *Saccharomyces cerevisiae* to improve multidimensional stress tolerance. *Applied and Environmental Microbiology,* **88**, e0162721.

299. Zhao, X. Q. and Bai, F. W. (2009). Mechanisms of yeast stress tolerance and its manipulation for efficient fuel ethanol production. *Journal of Biotechnology,* **144**, 23-30.

300. Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. and Futcher, B. (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular biology of the cell,* **9**, 3273-3297.

301. Breeden, L. L. (1997). Alpha-factor synchronization of budding yeast. *Methods in Enzymology,* **283**, 332-341.

302. Gómez-Ríos, D., Ramírez-Malule, H., Neubauer, P., Junne, S., Ríos-Estepa, R. and Ochoa, S. (2022). Tuning of fed-batch cultivation of *Streptomyces clavuligerus* for enhanced Clavulanic Acid production based on genome-scale dynamic modeling. *Biochemical Engineering Journal,* **185**, 108534.

Appendix A

Table Appendix A.1: All amino acids, nucleotide bases and reagents used in this thesis.

Amino acids/Nucleotide bases/Reagents	**Stock concentration**	**Final concentration**
Histidine	10 mg/mL in sterile water (500x)	0.02 mg/mL (1x)
Leucine	10 mg/mL in sterile water (100x)	0.1 mg/mL (1x)
Lysine	10 mg/mL in sterile water (333x)	0.03 mg/mL (1x)
Methionine	10 mg/mL in sterile water (500x)	0.02 mg/mL (1x)
Uracil	2 mg/mL in sterile water (100x)	0.02 mg/mL (1x)
Carbenicillin	100 mg/mL in sterile water (1000x)	100 µg/mL (1x)
G418	200 mg/mL in sterile water (1000x)	200 µg/mL (1x)
5-FOA	100 mg/mL in DMSO	Various, as specified in the thesis
D-Histidine	50 mg/mL in YNB	Various, as specified in the thesis
Cysteine	12.1 mg/mL in sterile water (100 mM)	0.0012 mg/mL (0.1 mM)

Appendix B

Table Appendix B.1: All relevant plasmids used and made in this thesis.

Plasmid	**Description**	**Use**	**Source**
pRS315	Low-copy, centromere yeast shuttle vector with LEU2 marker	Backbone for cloning	ATCC
pRS315-*wtpMET17-mCherry-tCYC1*	*mCherry* reporter under wild-type *MET17* promoter, terminated by *tCYC1*	For *MET17* promoter activity assay	This study
pRS315-mu*pMET17-mCherry-tCYC1*	*mCherry* reporter under the mutated *MET17* promoter, terminated by *tCYC1*	For *MET17* promoter activity assay	This study
pRS315-wild-type *FRT* site-KanMX- wild-type *FRT* site	*KanMX6* cassette flanked by wild-type *FRT* sites in the same orientation	For FLP/*FRT*-mediated recombination efficiency assay	This study

pRS315-High-3_sp1 *FRT* site-KanMX- High-3_sp1 *FRT* site	*KanMX6* cassette flanked by High-3_sp1 *FRT* sites in the same orientation	For FLP/*FRT*-mediated recombination efficiency assay	This study
pRS315-High-3_sp2 *FRT* site-KanMX-High-3_sp2 *FRT* site	*KanMX6* cassette flanked by High-3_sp2 *FRT* sites in the same orientation	For FLP/*FRT*-mediated recombination efficiency assay	This study
pRS315-High-3_sp4 *FRT* site-KanMX- High-3_sp4 *FRT* site	*KanMX6* cassette flanked by High-3_sp4 *FRT* sites in the same orientation	For FLP/*FRT*-mediated recombination efficiency assay	This study
pRS315-High-3_sp6 *FRT* site-KanMX- High-3_sp6 *FRT* site	*KanMX6* cassette flanked by High-3_sp6 *FRT* sites in the same orientation	For FLP/*FRT*-mediated recombination efficiency assay	This study
pRS315-*UP500(ura3)-1xUPRE1-SM-yeGFP-t*ADH1-*pURA3-URA3-tCYC1-DW500(ura3)*	UPR biosensor (*1xUPRE1-SM-yeGFP-t*ADH1) flanked by 500 bp homology regions for integration into *URA3* locus	For construction of UPR biosensor into *URA3* locus	This study

pRS315-*Up500(gsh1)-URA3-GFP-tADH1-Dw500(gsh1)*	*URA3-GFP-tADH1* flanked by 500 bp *GSH1* homology regions	For swapping *GSH1* gene with *URA3* marker	This study
pRS315- *Up500*(*pgsh1*)-*His3MX6-pTEF1-Dw500*(*pgsh1*)	*His3MX6-pTEF1* flanked by 500 bp *pGSH1* promoter homology regions	For replacing *pGSH1* promoter with *pTEF1* promoter	This study
pRS315-*Up500*(ura)-*pGSH1-URA3-GFP-tADH1-Dw500*(ura)	*pGSH1-URA3-GFP-tADH1* flanked by 500 bp *URA3* homology regions	For integration of *pGSH1-URA3-GFP-tADH1* into *URA3* locus	This study
pRS315-*Up500(GAP1)-KanMX-pGSH1-Dw500(GAP1)*	*KanMX-pGSH1* flanked by 500 bp *GAP1* promoter homology regions	For replacing *pGAP1* promoter with *pGSH1* promoter	This study
pCM190	Contains *URA3* marker and *CYC1* terminator	For amplification of *URA3* marker and *tCYC1*	Euroscarf
pFA6a-*HIS3MX6-pGAL1*	Contains *HIS3MX6* marker and *pGAL1* promoter	For construction of engineered BY4741 strain overexpressing *FLP* recombinase	M.S. Longtine, Univ. N. Carolina

pFA6a-*HIS3MX6-pGAL1-FLP-tCYC1*	*FLP* recombinase under *pGAL1*, terminated by *tCYC1*	For construction of engineered BY4741 strain overexpressing *FLP* recombinase	This study
pFA6a-GFP(S65T)-His3MX6	Contains GFP (S65T) and *His3MX6* marker	For amplification of GFP for C-terminal tagging and His3MX6 marker for selection	P. Philippsen, Univ. Basel
pFA6A-KanMX6	Contains the KanMX6 selectable marker	For amplification of the KanMX6 cassette	M.S. Longtine, Univ. N. Carolina
pSVA12	Contains *yeGFP-tADH1* fragment	For amplification of *yeGFP-tADH1* fragment	Avery's Lab
2-micron-based production plasmid for expressing SARS-CoV-2-mCherry or AP205 VLP-mCherry	See Figure 3.1 for construct details	For production of SARS-CoV-2-mCherry or AP205 VLP-mCherry	Phenotypeca Ltd.
pHR1-k	Empty 2-micron-based production plasmid vector with *LEU2* marker	Empty vector control of High-2 and High-3 strains	Phenotypeca Ltd.

Table Appendix B.2: All relevant oligos used in this thesis.

Oligo Name	Oligo Sequence 5' (overlap or overhang/ANNEAL) 3'	Use
***MET17* promoter activity assay**		
pMET17_F	taccgggccccccctcgaggGGATCTTCGGATGCAAGG	Amplification of wild-type or mutated *MET17* promoter
pMET17_R	tggagaccatTGTATGGATGGGGGTAATAG	
mCherry_F	catccatacaATGGTCTCCAAAGGTGAAGAAG	Amplification of *mCherry* ORF
mCherry_R	aattacatgaTTATTTGTACAATTCGTCCATAC	
tCYC1_F1	gtacaaataaTCATGTAATTAGTTATGTCACG	Amplification of *tCYC1*
tCYC1_R1	ctatagggcgaattggagctGCAAATTAAAGCCTTCGAG	
Quantification of plasmid copy number by ddPCR		
Plasmid_F	CACTAGTGCAGGATCAGGCC	Amplification of a specific non-coding region of the production plasmid for ddPCR
Plasmid_R	GCAGCACTTCCAACCTATGG	
ACT1_F	TGCAAACCGCTGCTCAATCT	Amplification of the *ACT1* gene in genomic DNA (reference for ddPCR)
ACT1_R	ACCGGCAGATTCCAAACCCA	
Construction of engineered *S. cerevisiae* BY4741 strain expressing FLP recombinase		

FLP_F	tgtctttaattaacccggggatccgATGCCACAATTTGATATATTATGTAAAAC	Amplification of *FLP* gene from the production plasmid
FLP_R	ctaattacatgaTTATATGCGTCTATTTATGTAGGATG	
tCYC1_F2	tagacgcatataaTCATGTAATTAGTTATGTCACGC	Amplification of *tCYC1*
tCYC1_R2	actatagaacgcggccgccagctgaGCAAATTAAAGCCTTCGAGC	
HO_int_F	cacacagggctatcagaaaatctataatatacagcaaaaaaccaaacacaCAGCAGTATAGCGACCAGCATTC	Amplification of *HIS3MX6-pGAL1-FLP-tCYC1* cassette for integration into *HO* locus
HO_int_R	ttttaaactttaaaattgtgacagctttccagaatggattatttttcctcGCAAATTAAAGCCTTCGAGCGTCC	
HO_int_CKF	GTTATGTGCGCAGATGGCTC	Check for correct integration
HO_int_CKR	GCGGACAGCATCAAACTGTAAG	
FLP/*FRT*-mediated recombination efficiency assay		
Ori_FRT_KAN_F	gggatcgtcgacgaagttcctattctctagaaagtataggaacttcGACATGGAGGCCCAGAATACC	Amplification of KANMX6 cassette with wild-type *FRT* sites in the same orientation

Ori_FRT_KAN_R	ccctaggagctcgaagttcctatactttctagagaataggaacttcCAGTATAGCGACCAGCATTCAC	
High-3_sp1_F	gggatcgtcgacgaagttcctattctctctagctagaaagtataggaacttcGACATGGAGGCCCAGAATACC	Amplification of KANMX6 cassette with High-3_sp1 *FRT* sites in the same orientation
High-3_sp1_R	ccctaggagctcgaagttcctatactttagaataggaacttcCAGTATAGCGACCAGCATTCAC	
High-3_sp2_F	gggatcgtcgacgaagttcctattctctctagaaagtataggaacttcGACATGGAGGCCCAGAATACC	Amplification of KANMX6 cassette with High-3_sp2 *FRT* sites in the same orientation
High-3_sp2_R	ccctaggagctcgaagttcctatactttctagagaataggaacttcCAGTATAGCGACCAGCATTCAC	
High-3_sp4_F	gggatcgtcgacgaagttcctattctctctagaaagtataggaacttcGACATGGAGGCCCAGAATACC	Amplification of KANMX6 cassette with High-3_sp4 *FRT* sites in the same orientation
High-3_sp4_R	ccctaggagctcgaagttcctatactttctagagagaataggaacttcCAGTATAGCGACCAGCATTCAC	
High-3_sp6_F	gggatcgtcgacgaagttcctattctctctagagaaagtataggaacttcGACATGGAGGCCCAGAATACC	Amplification of KANMX6 cassette with High-3_sp6 *FRT* sites in the same orientation

High-3_sp6_R	ccctaggagctcgaagttcctatactttagaataggaacttcCAGTATAG CGACCAGCATTCAC	
***MMS21* locus sequencing**		
*MMS21*_F	CCTCTTTATCCATCTATGCG	Colony PCR and sequencing of the *MMS21* locus to analyse natural variation
MMS21_R	CCGGTATTACACAAGTTGG	
Construction of UPR reporter system		
1x*UPRE1-SM*_F	tttaatttgcGGACAGCGTGTCCTTAAGATCTTGTAATATT CTAATCAAGCTTATAAAAGAGCACTGTTGGGCGTGA GTGGAGGCGCCGGAATAAACCATGCAATGAA	Annealing of the forward and reverse oligos for generating the 1x*UPRE1-SM* UPR promoter sensor
1x*UPRE1-SM*_R	ctttagacatTTCATTGCATGGTTTATTCCGGCGCCTCCA CTCACGCCCAACAGTGCTCTTTTATAAGCTTGATTA GAATATTACAAGATCTTAAGGACACGCTGTCC	
*UP_ura3*_F	taccgggccccccctcgaggGAAACTAGGGAAGACAAGC	Amplification of homology arm corresponding to the upstream of the *URA3 ORF*
*UP_ura3*_R	ttgaattgaaTGGACCCTGAAACCACAG	
*pURA3-URA3*_F	tcagggtccaTTCAATTCAATTCATCATTTTTTTTTTATTC TTTTTTTTG	Amplification of *URA3* selectable marker under control of its native promoter

*pURA3-URA3*_R	aattacatgaTTAGTTTTGCTGGCCGCATC	
tCYC1_F	gcaaaactaaTCATGTAATTAGTTATGTCACG	Amplification of *CYC1* terminator
tCYC1_R	cacgctgtccGCAAATTAAAGCCTTCGAG	
*yeGFP-t*ADH1_F	atgcaatgaaATGTCTAAAGGTGAAGAATTATTCACTG	Amplification of the yeast GFP reporter sequence *yeGFP-t*ADH1
*yeGFP-t*ADH1_R	gagattcccgCCGGTAGAGGTGTGGTCAATAAG	
*DW_ura3*_F	cctctaccggCGGGAATCTCGGTCGTAATG	Amplification of homology arm corresponding to the downstream of the *URA3 ORF*
*DW_ura3*_R	ctatagggcgaattggagctGGGCTACACCAGAGATAC	
*URA_INT*_F	GAAACTAGGGAAGACAAGCAAC	*URA_INT_F* and *URA_INT_R* were used for amplification of the *UP_ura3-1xUPRE1-SM-yeGFP-t*ADH1-*pURA3-URA3-tCYC1-DW_ura3* cassette for integration into *URA3* locus; *URA_CKF* was used with *URA_INT_R* to confirm correct integration at *URA3* locus
*URA_INT*_R	GGGCTACACCAGAGATAC	
*URA*_CKF	CTACCGTTGCTAAATTCGAGTG	
Construction of a diploid heterozygous yeast strain incorporating a *URA3*-based feedback selection system		
*UP_pgsh1*_F	taccgggccccccctcgaggCTCATGCTTTGGTTGGCTTC	Amplification of homology arm corresponding to the upstream of the *pGSH1* promoter
*UP_pgsh1*_R	cctccatgtcGTGTGTGTGCGTAAGCGTATG	

*HIS3MX6*_F	gcacacacacGACATGGAGGCCCAGAATAC	Amplification of *HIS3MX6* selection marker
*HIS3MX6*_R	tggtgtgtggCAGTATAGCGACCAGCATTC	
*pTEF1*_F	cgctatactgCCACACACCATAGCTTCAAAATG	Amplification of *pTEF1* promoter
*pTEF1*_R	gtcccattttCTTAGATTAGATTGCTATGCTTTCTTTC	
*DW_pgsh1*_F	ctaatctaagAAAATGGGACTCTTAGCTTTG	Amplification of homology arm corresponding to the downstream of the *pGSH1* promoter
*DW_pgsh1*_R	ctatagggcgaattggagctCCTCAACGTAACTACCCAC	
*pGSH1_SWAP*_F	CTCATGCTTTGGTTGGCTTC	Amplification of the *Up_pgsh1-His3MX6-pTEF1-Dw_pgsh1* cassette for replacement of *pGSH1* with *pTEF1*
*pGSH1_SWAP*_R	CCTCAACGTAACTACCCAC	
*UP_gsh1*_F	taccgggccccccctcgaggGTGACTACAACTGTGGCTG	Amplification of homology arm corresponding to the upstream of the *GSH1* ORF
*UP_gsh1*_R	ctttcgacatTTTATTCTTCTATATGTATATTTTCGATACTC	
*URA3*_F	gaagaataaaATGTCGAAAGCTACATATAAGGAACG	Amplification of *URA3* ORF
*URA3*_R	cggggatccgGTTTTGCTGGCCGCATCTTC	
*GFP*_F	ccagcaaaacCGGATCCCCGGGTTAATTAAC	Amplification of C-terminal fused *GFP* ORF
*GFP*_R	taaaaggagtCTATTTGTATAGTTCATCCATGCC	

*DW_gsh1*_F	atacaaatagACTCCTTTTACTTCGGTTG	Amplification of homology arm corresponding to the downstream of the *GSH1* ORF
*DW_gsh1*_R	ctatagggcgaattggagctCATACCTTGCCCTCAGAG	
*URA3*_INT_F	GTGACTACAACTGTGGCTG	Amplification of the *Up_gsh1-URA3-GFP-tADH1-Dw_gsh1* cassette for replacement of *GSH1* ORF with *URA3* ORF fused with *GFP*
*URA3*_INT_R	CATACCTTGCCCTCAGAG	
*GSH1locus*_CKF	AACTCTCCGTAACCACCAAG	Confirm correct replacement of *pGSH1* with *pTEF1;* Confirm correct replacement of *GSH1 ORF* with *URA3-GFP*; used for PCR analysis of *GSH1* locus
*GSH1locus*_CKR	GCCAATATTCTACGTGCTTG	
Construction of a diploid homozygous yeast strain incorporating a *URA3*-based feedback selection system		
*UP_ura3*_F	taccgggccccccctcgaggGAAACTAGGGAAGACAAGC	Amplification of homology arm corresponding to the upstream of the *ura3* locus
*UP_ura3*_R	gcgtgtgcacTGGACCCTGAAACCACAG	
*pGSH1*_F	tcagggtccaGTGCACACGCCTGTTACTTC	Amplification of *pGSH1* promoter
*pGSH1*_R	ctttcgacatTTTATTCTTCTATATGTATATTTTCGATACT CTAAACC	

*URA3-GFP-tGSH1*_F	gaagaataaaATGTCGAAAGCTACATATAAG	Amplification of *URA3-GFP-tGSH1* cassette from the previous engineered diploid heterozygous yeast genomic DNA
*URA3-GFP-tGSH1*_R	gagattcccgTATCGAATAGTGTGAACCG	
*DW_ura3*_F	ctattcgataCGGGAATCTCGGTCGTAATG	Amplification of homology arm corresponding to the downstream of the *ura3* locus
*DW_ura3*_R	ctatagggcgaattggagctGGGCTACACCAGAGATAC	
*URA_INT*_F	GAAACTAGGGAAGACAAGCAAC	*URA_INT*_F and *URA_INT*_R were used for amplification of the *UP_ura3-pGSH1_URA3-GFP-tGSH1-DW_ura3* cassette for integration into *URA3* locus; *URA_CKF* was used with *URA_INT_R* to confirm correct integration at *URA3* locus
*URA_INT*_R	GGGCTACACCAGAGATAC	
*URA*_CKF	CTACCGTTGCTAAATTCGAGTG	
Construction of the control *pURA3-URA3* yeast strain		
*UP_ura3*_F	taccgggccccccctcgaggGAAACTAGGGAAGACAAGC	Amplification of homology arm corresponding to the upstream of the *ura3* locus
*UP_ura3*_R	ttgaattgaaTGGACCCTGAAACCACAG	
*pURA3*_F	tcagggtccaTTCAATTCAATTCATCATTTTTTTTTTATTC	Amplification of *pURA3* promoter

*pURA3*_R	ctttcgacatGATTTATCTTCGTTTCCTGC	
*URA3-GFP-tGSH1*_F	aagataaatcATGTCGAAAGCTACATATAAG	Amplification of *URA3-GFP-tGSH1* cassette from the previous engineered diploid heterozygous yeast genomic DNA
*URA3-GFP-tGSH1*_R	gagattcccgTATCGAATAGTGTGAACCG	
*DW_ura3*_F	ctattcgataCGGGAATCTCGGTCGTAATG	Amplification of homology arm corresponding to the downstream of the *ura3* locus
*DW_ura3*_R	ctatagggcgaattggagctGGGCTACACCAGAGATAC	
*URA_INT*_F	GAAACTAGGGAAGACAAGCAAC	*URA_INT*_F and *URA_INT*_R were used for amplification of the *UP_ura3-pURA3_URA3-GFP-tGSH1-DW_ura3* cassette for integration into *URA3* locus; *URA_CKF* was used with *URA_INT_R* to confirm correct integration at *URA3* locus
*URA_INT*_R	GGGCTACACCAGAGATAC	
*URA*_CKF	CTACCGTTGCTAAATTCGAGTG	
Construction of a diploid homozygous yeast strain incorporating a *GAP1*-based feedback selection system		
*UP_gap1*_F	taccgggccccccctcgaggCCATCAGATGGTGTCTTG	Amplification of homology arm corresponding to the upstream of the *GAP1* locus
*UP_gap1*_R	cctccatgtcGGTCAATGGGTAGATTAG	

*KANMX*_F	cccattgaccGACATGGAGGCCCAGAATAC	Amplification of *KANMX6* cassette for selection
*KANMX*_R	gcgtgtgcacCAGTATAGCGACCAGCATTC	
*pGSH1*_F	cgctatactgGTGCACACGCCTGTTACTTC	Amplification of *pGSH1* promoter
*pGSH1*_R	tattactcatTTTATTCTTCTATATGTATATTTTCGATACTC TAAACC	
*DW_gap1*_F	gaagaataaaATGAGTAATACTTCTTCGTAC	Amplification of homology arm corresponding to the downstream of the *GAP1* locus
*DW_gap1*_R	ctatagggcgaattggagctCGTAAATCATGGTACCTG	
*GAP1_INT*_F	CCATCAGATGGTGTCTTG	*GAP1_INT*_F and *GAP1_INT*_R were used for amplification of the *UP_gap1-KANMX6-pGSH1-DW_gap1* cassette for replacement of *pGAP1* with *pGSH1*; *GAP1_CKF* was used with *GAP1_INT_R* to confirm correct replacement of *pGAP1*
*GAP1_INT*_R	CGTAAATCATGGTACCTG	
*GAP1*_CKF	GCCAAGGACAATATGTTGCG	
Construction of the control *pCYC1-GAP1* yeast strain		
*UP_gap1*_F	taccgggccccccctcgaggCCATCAGATGGTGTCTTG	Amplification of homology arm corresponding to the upstream of the *GAP1* locus
*UP_gap1*_R	cctccatgtcGGTCAATGGGTAGATTAG	

*KANMX*_F	cccattgaccGACATGGAGGCCCAGAATAC	Amplification of *KANMX6* cassette for selection
*KANMX*_R	ccaaatgaacCAGTATAGCGACCAGCATTC	
*pCYC1*_F	cgctatactgGTTCATTTGGCGAGCGTTG	Amplification of *pCYC1* promoter
*pCYC1*_R	agtattactcatTATTAATTTAGTGTGTGTATTTGTGTTTGTG	
*DW_gap1*_F	taaattaataATGAGTAATACTTCTTCGTAC	Amplification of homology arm corresponding to the downstream of the *GAP1* locus
*DW_gap1*_R	ctatagggcgaattggagctCGTAAATCATGGTACCTG	
*GAP1_INT*_F	CCATCAGATGGTGTCTTG	*GAP1_INT*_F and *GAP1_INT*_R were used for amplification of the *UP_gap1-KANMX6-pCYC1-DW_gap1* cassette for replacement of *pGAP1* with *pCYC1*; *GAP1_CKF* was used with *GAP1_INT_R* to confirm correct replacement of *pGAP1*
*GAP1_INT*_R	CGTAAATCATGGTACCTG	
*GAP1*_CKF	GCCAAGGACAATATGTTGCG	
RT-qPCR primers for GSH-related genes		
Q_*GAP1*_F	GGTCCCGCTTCGCTACTGAT	Quantification of *GAP1* expression
Q_*GAP1*_R	ACAGCCAACTCACCCAGAGC	

Q_*GSH1*_F	ACGCTGACTGTCTTCCCGAG	Quantification of *GSH1* expression
Q_*GSH1*_R	GCCTGGTCCTGATGGATGCT	
Q_*GSH2*_F	GAGCAGGAAATAGCGGTGGT	Quantification of *GSH2* expression
Q_*GSH2*_R	GCCTCCCAGTCCTTTTCGGA	
Q_*ACT1*_F	TGCAAACCGCTGCTCAATCT	Quantification of *ACT1* reference gene expression
Q_*ACT1*_R	ACCGGCAGATTCCAAACCCA	

Appendix C

Table Appendix C.1: Specific sequence changes identified by whole plasmid sequencing.

Types of mutation (description)	**Subtypes of mutations (description)**	**10th generation plasmid sequence**	**140th generation plasmid sequence**
Type 1 (at *FRT* sites)	Type 1a (at *FRT* site upstream of the FLP; four nt addition) spacer from 8bp to 12 bp	TCTAG----AAA	TCTAGCTAGAAA
	Type 1b (at *FRT* site upstream of the FLP; two nt addition) spacer from 8bp to 10 bp	TCT--AGAAA	TCTCTAGAAA
	Type 1c (at *FRT* site downstream of the FLP; four nt deletion) spacer from 8bp to 4 bp	TTTCTAGA	TTT----A
	Type 1d (at *FRT* site downstream of the FLP; two nt addition) spacer from 8bp to 10 bp	TTTCTAGA--GAA	TTTCTAGAGAGAA
	Type 1e (at *FRT* site upstream of the FLP; five nt addition) spacer from 8bp to 13bp spacer	TCT-----AGAAA	TCTCTAGCTAGAAA
	Type 1f (at *FRT* site downstream of the FLP; two nt deletion) spacer from 8bp to 6bp	TTTCTAGA	TTT--AGA

Type 2 (at SARS-CoV-2 spike protein fragment sequence)	Type 2a (four substitution, one addition)	CAATCCTACGG CTT-TCAACC	CAATCCTACGTCCATGTAACC
	Type 2b (four deletion, five addition)	CAATCCTACGG CTTTC-----AACC	CAATCCTACG---- TCCATGTAACC
Type 3 (at pMET17 sequence)		CACGTGAAGCT GTCGATA	CACGTAAAGCTGTCGATA
Type 4 (at mCherry sequence; nonsense mutation)		ATGGTCTCCAA AGGTGAAG	ATGGTCTCCAAAGGTTAAG
Type 5 (at terminator region of the expression cassette; downstream of mcherry sequence; big chunk deletion)		ACACCTCTACC GGCATGCCGA GCAAATGCCTG CAAATCGCTCC CCATTTCACCC AATTGTAGATAT GCTAACTCCAG	ACAC----ACCG-CATCAC---- AAATGCCT-CAAATCGCT------- ---------- ---------- GTAA

		CAATGAGTTGA TGAATCTCGGT GTGTATTTTATG TCCTCAGAGGA CAACACCTGTT GTAA	
Type 6 (at FLP region; a part of FLP seq duplicated)		TCCCAGATACT TATTTTGGACTA ATTT-	TCCCAGATACTTATTTTGGAC TAATTTAAATGATTTCGGATC AACGTTCTTAATATCGCTGAA TCTTCCACAATTGATGAAAGT AGCTAGGAAGAGGAATTGGT ATAAAGTTTTTGTTTTTGTAAA TCTCGAAGTATACTCAAACGA ATTTAGTATTTTCTCAGTGAT CTCCCAGATGCTTTCACCCTC ACTTAGAAGTGCTTTAAGCAT TTTTTTACTGTGGCTATTTCC CTTATCTGCTTCTTCCGATGA TTCGAACTGTAATTGCAAACT ACTTACAATATCAGTGATATT TTGGACTAATTT
Type 7 (at LEU2 promoter region)		GAGAACA----------	GAGAACATAACGAGAACACA CAGGGGCGCTATCGCAC

		CACAGGGGCG CTATCGCAC	
Type 8 (at noncoding regions)	Type 8a (at the noncoding region upstream of the expression cassette; one nt deletion)	TAAAAAAAAAAA AATAC	TAAAAAAAAAAAA-TAC
	Type 8b (at the noncoding region upstream of the expression cassette; one nt addition)	TAAAAAAAAAAA AA-TAC	TAAAAAAAAAAAAAATAC
	Type 8c (at the noncoding region downstream of the expression cassette; one nt deletion)	TTTTCGTATTG	TTTT-GTATTG

Appendix D

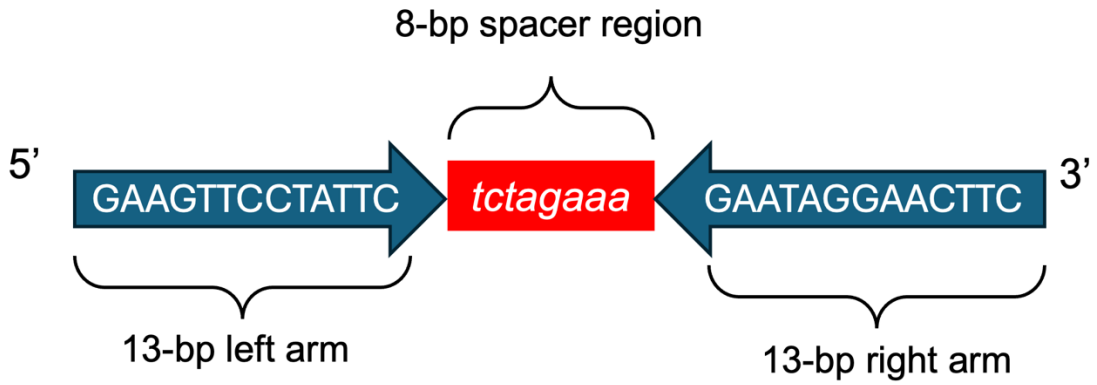


Figure Appendix D.1: Illustration of the minimal *FRT* site sequence. The minimal *FRT* site consists of two 13-bp inverted repeat arms flanking an 8-bp spacer region. The left arm and right arm are binding sites for the FLP recombinase, while the central 8-bp spacer region determines the site of strand exchange and FLP recombinase cleavage during recombination.