CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	Strategies to design bioprocesses for sustainable yeast-based production of recombinant proteins and bioproducts
Authors	Belloch-Molina, Carlos
Publication date	2025-12-30
Original Citation	Belloch-Molina, C. 2025. Strategies to design bioprocesses for sustainable yeast-based production of recombinant proteins and bioproducts. PhD Thesis, University College Cork.
Type of publication	Doctoral thesis
Rights	© 2025, Carlos Belloch Molina. - https://creativecommons.org/licenses/by-nc/4.0/
Download date	2026-11-17 11:49:39
Item downloaded from	https://hdl.handle.net/10468/19269

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Ollscoil na hÉireann, Corcaigh

National University of Ireland, Cork



Strategies to design bioprocesses for sustainable yeast-based production of recombinant proteins and bioproducts

Thesis presented by

Carlos Belloch Molina, MSc

for the degree of

Doctor of Philosophy

University College Cork

School of Microbiology & School of Engineering and Architecture

Head of School/Department: Prof. David Clarke

Supervisors: Prof. John Morrissey & Prof. Maria de Sousa Gallagher

2025

Table of contents

Declaration of integrity.....2

Abstract.....3

List of abbreviations.....4

Manuscripts and publications.....6

General introduction.....7

Chapter 1.....11
Engineering nature: The technological cornerstones of modern synthetic biology

Chapter 2.....69
Unravelling *Kluyveromyces marxianus*' nitrogen needs for improved dairy biorefineries

Chapter 3.....120
MoClo-compatible secretion expansion and Yeast Surface Display platform for *Kluyveromyces marxianus*

Chapter 4.....202
Evolving an improved strain of *Kluyveromyces marxianus* for use in a dairy biorefinery

Chapter 5.....244
Insights and future perspectives

Acknowledgements.....256

Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

Signed……. Carlos Belloch Molina

Abstract

Kluyveromyces marxianus is a fast-growing, thermotolerant yeast with the ability to metabolise diverse sugars and naturally secrete proteins, making it a promising candidate for sustainable biorefineries. However, its industrial adoption has been limited by a lack of standardised genetic tools and incomplete understanding of its complex regulatory networks. This thesis addresses these bottlenecks by integrating transcriptomics, synthetic biology, and Adaptive Laboratory Evolution to characterise, engineer, and evolve *K. marxianus* into a robust and versatile industrial chassis.

First, transcriptome analysis was used to elucidate the nitrogen requirements of *K. marxianus* during growth on whey permeate, revealing activation of nitrogen scavenging pathways and unique adaptations. These insights provide a foundation for optimising growth and productivity in dairy-based waste streams. Next, a modular, Golden Gate compatible toolkit was developed to expand protein secretion capabilities. Native secretion signals were functionally characterised using a Yeast Surface Display platform, enabling standardised study of heterologous protein secretion in *K. marxianus*. Finally, Adaptive Laboratory Evolution was applied to generate a non-Genetically Modified strain with enhanced ethanol tolerance and operational robustness. This evolved strain was validated both in laboratory-scale fermentations and, most importantly, in full-scale trials at the Carbery Food Ingredients plant, demonstrating its industrial applicability for converting whey permeate into bioethanol.

Together, this work delivers functional insights, practical synthetic biology tools, and industrial validation, advancing *K. marxianus* as a resilient microbial cell factory for circular dairy biorefineries. By bridging these fields, this thesis establishes a foundation for future deployment of *K. marxianus* as a cornerstone organism in the bioeconomy.

Abbreviations

A_{600}	Absorbance at 600 nm wavelength
AFM	Alcohol Fermentation Monitor
AI	Artificial Intelligence
ALE	Adaptive Laboratory Evolution
BCAA	Branched Chain Amino Acid
BOD	Biological Oxygen Demand
bp	base pair
CDS	Coding Sequence
CDW	Cell Dry Weight
CHP	Combined Heat and Power
CPP	Critical Process Parameters
CQA	Critical Quality Attributes
CV	Coefficient of Variation
DBTL	Design-Build-Test-Learn
DGE	Differential Gene Expression
DoE	Design of Experiments
EFSA	European Food Safety Authority
ER	Endoplasmic Reticulum
EU	European Union
EvA	Evolved strain A
EvB	Evolved strain B
EvC	Evolved strain C
EvD	Evolved strain D
FDA	U.S. Food and Drug Administration
GABA	gamma-aminobutyrate
GMO	Genetically Modified Organism
GO	Gene Ontology
GPI	GlycosylPhosphatidylInositol
GRAS	Generally Recognized As Safe
GRAVY	Hydrophobicity index
gRNA	guide RNA
HA	Hemagglutinin Antigen
HPLC	High-Performance Liquid Chromatograph
HSD	Honestly Significant Difference
iGEM	International Genetically Engineered Machine
IP	Isoelectric Point
LLM	Large Language Models
MMMel	Minimal Media Melibiose
MoClo	Modular Cloning
N-	Nitrogen
ONT	Oxford Nanopore Technologies
PacBio	Pacific Biosciences
PCA	Principal Component Analysis

PIR	Protein with Internal Repeats
PNPG	p-nitrophenyl-α-d-galactopyranoside
POI	Protein of Interest
QbD	Quality by Design
QPS	Qualified Presumption Safety
RF	Repair Fragment
RFU	Relative Fluorescence units
RIN	RNA Integrity Number
RO	Reverse Osmosis
RT	Room Temperature
SAM	S-adenosyl-methionine
SCRaMbLE	Synthetic Chromosome Rearrangement and Modification by LoxP-mediated Evolution
SDG	Sustainable Development Goals
SS	Secretion Signal
SSF	Simultaneous Saccharification and Fermentation
TAR	Transformation-Associated Recombination
UN	United Nations
WGD	Whole Genome Duplication
WP	Whey Permeate
WPT	Whey Permeate Tryptone
WT	Wild Type
YPD	Yeast extract Peptone Dextrose
YPLac	Yeast extract Peptone Lactose
YSD	Yeast Surface Display
YTK	Yeast Tool Kit
αMF	α Mating Factor
μ_{max}	Maximum specific growth rate

Publications in preparation for submission

Belloch-Molina C, Morrissey JP and de Sousa Gallagher M. Engineering nature: The technological cornerstones of modern synthetic biology.

Belloch-Molina C, Fenton DA, de Sousa Gallagher M and Morrissey JP. Unravelling *Kluyveromyces marxianus*' nitrogen needs for improved dairy biorefineries.

Belloch-Molina C, Romano R, Kishkevich A, Ellis T, de Sousa Gallagher M and Morrissey JP. MoClo-compatible secretion expansion and Yeast Surface Display platform for *Kluyveromyces marxianus*.

Belloch-Molina C, Donzella L, Linguardo F, Walsh C, O'Neill A, Mulcahy P, Lyons M, Hallihan A, de Sousa Gallagher M and Morrissey JP. Evolving an improved strain of *Kluyveromyces marxianus* for use in a dairy biorefinery.

Other publications during this Thesis

Flavin S, Meltonville M, Taverner C, Reid J, Lawrence S, Belloch-Molina C & Morrissey J (2023). Understanding Early Modern Beer: An Interdisciplinary Case-Study. The Historical Journal, 66(3), 516–549. doi:10.1017/S0018246X23000043

Donzella L, Belloch-Molina C, Morrissey JP and Sousa MJ. Reprogramming a Hgt-family *Kluyveromyces marxianus* sugar transporter to enable co-consumption of glucose and xylose.

General introduction

1. The need for a circular bioeconomy

The global industrial landscape is undergoing a paradigm shift, driven by the urgent necessity to decouple economic growth from fossil resources consumption. The current climate crisis, coupled with the depletion of finite resources, has catalysed the transition from a linear economic model to a circular bioeconomy. At the heart of this transition lies the concept of the **Biorefinery**. Much like their petrochemical counterparts, biorefineries aim to convert raw materials into a spectrum of value-added products, fuels, and chemicals. However, unlike oil refineries, they rely on renewable biomass and industrial side-streams as feedstocks, offering a sustainable path to decarbonisation.

2. Microbial cell factories for the biorefinery

The success of a biorefinery is intrinsically linked to the efficiency of the biological platforms, or microbial cell factories, that power them. Historically, the baker's yeast *Saccharomyces cerevisiae* has served as the premier workhorse for industrial biotechnology. Its robust fermentation capabilities and well-characterised genome have cemented its status as a model organism. However, the diverse demands of modern biorefineries often exceed the physiological limits of *S. cerevisiae*. High-temperature processes, the utilisation of complex industrial side-streams (such as whey permeate), and the efficient secretion of proteins often require capabilities that conventional *S. cerevisiae* struggles to provide without extensive and often metabolically burdensome engineering.

3. The rise of non-conventional yeasts

To overcome these limitations, industrial biotechnology is increasingly turning towards alternative or so-called "non-conventional" yeasts. Among these, *K. marxianus* stands out as an exceptional candidate. Often described as the fastest-growing eukaryote, *K. marxianus* possesses innate traits that are highly desirable for heterologous industrial applications: thermotolerance, the ability to utilise a broad

range of substrates, including lactose and pentose sugars, and a metabolism better suited for protein production.

While the shift from *S. cerevisiae* towards less domesticated yeasts such as *K. marxianus* is relatively recent, the landscape has shifted dramatically in the last decade. The field has moved beyond simple random integration methods, witnessing a surge in the development of sophisticated molecular resources. These include high-quality genome assemblies and annotations that facilitate omics-level understanding, CRISPR-Cas9 toolkits for precise genome editing, and Modular Cloning (MoClo) libraries that enable rapid pathway assembly. However, despite this toolbox, gaps remain in harnessing specific physiological traits, for industrial-scale applications.

4. Thesis Objectives and Scope

The research carried out during my doctoral period and reported in this thesis aims to contribute to bridging the gap between the potential of *K. marxianus* and its industrial implementation. By integrating diverse methodologies, ranging from transcriptomics and synthetic biology to Adaptive Laboratory Evolution (ALE), this work seeks to characterise, engineer, and evolve *K. marxianus* into a superior chassis for sustainable biorefineries.

The research presented herein addresses three critical bottlenecks:

1. **Understanding:** Deciphering the metabolic requirements of the host. In particular the nitrogen metabolism of *K. marxianus* is extrapolated from *S. cerevisiae*. However, key differences exist, which support the need for a deeper, species-specific investigation

2. **Engineering:** Developing complex and standardised tools for protein production. This work focuses on establishing a robust secretion toolkit to unlock the potential of *K. marxianus* as a cell factory for secretion of high-value proteins.

3. **Application:** Validating industrial robustness and scalability. Using ALE to overcome physiological limitations in harsh industrial environments,

demonstrating the viability of this chassis through a large-scale dairy side-stream valorisation case study.

5. Outline of the Thesis

The thesis is organised into five chapters, moving from the state-of-the-art in synthetic biology to specific applications in *K. marxianus*:

Chapter 1. Engineering nature: The technological cornerstones of modern synthetic biology establishes the theoretical framework. This chapter reviews the evolution of synthetic biology and explores how concepts imported from engineering, like the "Design-Build-Test-Learn" cycle, are revolutionising our ability to set up sustainable biorefineries.

Chapter 2. Unravelling *Kluyveromyces marxianus'* nitrogen needs for improved biorefineries dives into the fundamental physiology of the organism. Nitrogen is a critical, often limiting nutrient in industrial fermentations. By analysing transcriptome data, this chapter investigates how *K. marxianus* senses and responds to nitrogen availability, providing crucial insights into metabolic bottlenecks that affect growth and productivity in the particular case of whey permeate.

Chapter 3. MoClo-Compatible Secretion Expansion and Yeast Surface Display Platform for *Kluyveromyces marxianus* moves from understanding to engineering. One of the major advantages of yeasts over bacteria is their ability to secrete complex proteins. This chapter details the development of a synthetic biology toolkit compatible with the Golden Gate standard. By isolating novel secretion signals and characterising them using Yeast Surface Display technology, this work provides the scientific community with a versatile set of parts for high-level protein secretion in *K. marxianus*.

Chapter 4. Evolving an improved strain of *Kluyveromyces marxianus* for use in a dairy biorefinery focuses on industrial application. While genetic engineering is powerful, regulatory constraints in the food industry often need non-GMO (Genetically Modified Organism) approaches. This chapter describes the development of a robust strain capable of valorising whey permeate into ethanol.

Through the use of ALE, a phenotype that outperforms wild-type strains in whey permeate was achieved, demonstrating successful scale-up to 120,000 L.

Finally, **Chapter 5: Insights and future perspectives**, synthesises the findings of this research, discussing how the combination of omics, synthetic biology tools, and evolutionary engineering can further consolidate *K. marxianus* as a valuable candidate for the establishment of bioprocesses for the future bioeconomy.

Chapter 1

Engineering nature: The technological cornerstones of modern synthetic biology

Abstract

The growing demand for sustainable alternatives to traditional manufacturing processes has increased interest in biotechnological solutions. Synthetic biology has emerged as a transformative discipline that applies engineering principles to the rational design of biological systems with predictable and tuneable behaviour. Supported by the Design-Build-Test-Learn framework and complementary methodologies such as Quality by Design, synthetic biology enables the systematic development of biological platforms for industrial, environmental, agricultural, and biomedical applications, including advanced biorefinery concepts. This review outlines the conceptual foundations of synthetic biology and its rapid expansion, driven by advances in DNA sequencing and synthesis, genome editing, high-throughput assembly methods, computational modelling, and artificial intelligence-driven design. It further discusses the role of synthetic biology in enabling a future circular bioeconomy, while highlighting the technical, scalability, regulatory, and societal challenges that must be addressed for successful translation beyond the laboratory.

Introduction

The 2030 Agenda for Sustainable Development, a global plan of action for sustainable development, adopted by the United Nations (UN) in 2015 contains at its core the Sustainable Development Goals (SDGs) (United Nations, 2015a), a set of 17 interconnected global goals adopted by all UN Member States. These goals serve as a universal call to action to end poverty, protect the planet, and ensure prosperity for all by 2030. However, the escalating environmental challenges have put under serious risk achieving on time these goals. Traditional manufacturing processes relying on finite resources not only contribute to environmental degradation but also aggravate climate change. Thus, there is urgent need for sustainable solutions that can alleviate climate related issues and prevent nature degradation (United Nations, n.d.-b). Nevertheless, although nature's diversity is heavily endangered, biological systems are naturally resilient and adaptive, having evolved over millions of years to survive in changing environments and still enduring within the new conditions we are imposing (Brans et al., 2017; Palumbi, 2001; Reid et al., 2016; Yoshida et al., 2016). Microorganisms possess inherent mechanisms to acclimatise to and even revert environmental stresses, and by harnessing nature's power to adapt, powerful tools are being developed to fight environmental problems. Synthetic biology, with its ability to engineer biological systems for specific purposes, holds immense potential for revolutionising biotech product development and manufacturing in a sustainable manner (Elowitz & Leibier, 2000; Endy, 2005; Huang et al., 2001; Khalil & Collins, 2010; Lorenzo et al., 2018). Biorefineries, are also a critical innovation, offering the workflow for converting renewable feedstocks into valuable products, further aligning with the principles of sustainability (Dahiya et al., 2018; Leong et al., 2021; Venkata Mohan et al., 2016). By channelling the synergies between synthetic biology and biorefineries, we can accelerate the transition towards a bio-based economy, where renewable resources are efficiently utilised to meet production demand while minimising environmental impact.

The Rise of Synthetic Biology

Nature holds immense potential for the development of sustainable and affordable solutions for industrial issues, questions and obstacles. However, the complexity of biological systems often makes it challenging to engineer and predict desired outputs in bioengineered entities. To address this challenge, different engineering concepts and ideas have been introduced for the development of tailored biological organisms: standardisation, modularity, orthogonality or abstraction are at the core of this transfer (Endy, 2005; Meng & Ellis, 2020). These concepts, together with the advances in DNA and protein technologies, led to the establishment of synthetic biology. This groundbreaking field offers unprecedented capabilities for designing and constructing biological systems with tailored functions, emerging as a transformative field at the intersection of molecular biology, engineering, and computational sciences.

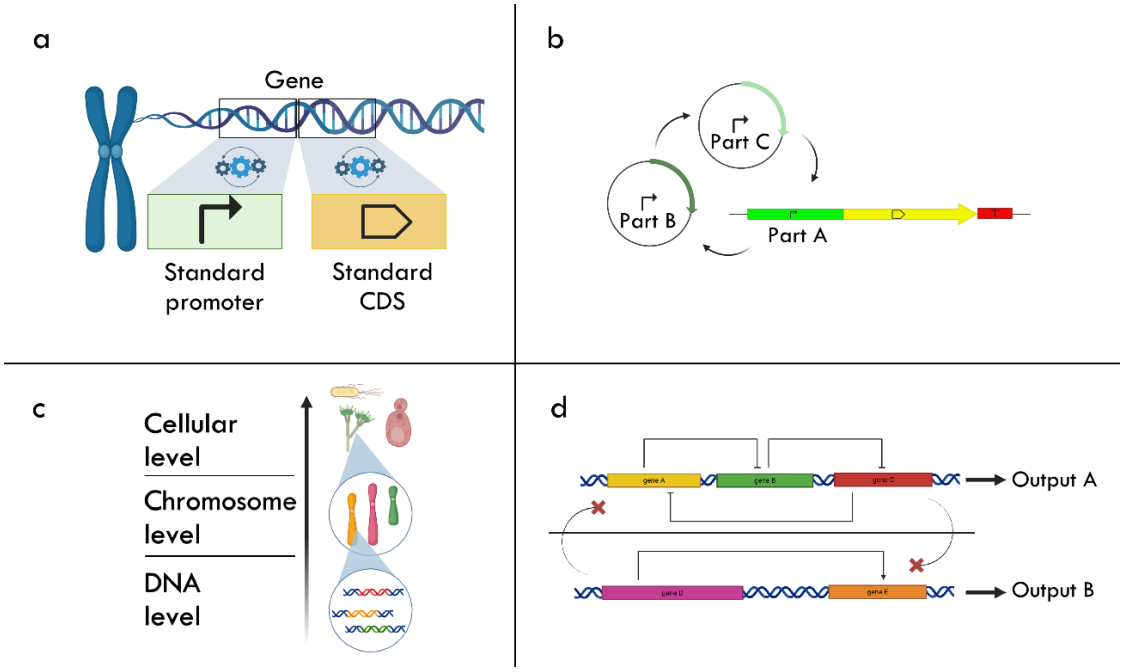


Fig. 1. Concepts imported from engineering fields and adapted to biology. **a)** Standardisation happens at the genetic level when gene parts such as promoters or coding sequences (CDS) are identified, extracted and adapted for combinatorial design, so that their output is defined. **b)** Modularity allows the substitution of standard parts for others that fulfil the same function, allowing interchangeability. **c)** Abstraction enables focusing on a particular level of complexity without considering challenges spilling from other levels of complexity. **d)** Orthogonality prevents cross-interactions between designed systems so that output is always the expected.

Standardisation (Fig. 1a), wherein complex systems are reduced to their minimum parts with well-defined functions (Canton et al., 2008; Galdzicki et al., 2014), is one of the central concepts imported from engineering. Once biological parts are isolated, they can then be made available in standardised formats to be characterised and allowing them to be combined into more complex systems with expected outputs following uniform procedures and methods, facilitating interchangeability of results. Standardisation spans different techniques, enabling the combination of these parts to design biological systems with desired functionalities.

A second idea imported from engineering is the concept of **modularity** (Fig. 1b), which allows for the decoupling of complicated systems into standard parts, enabling the construction of more complex genetic circuits from standardised components (Butti et al., 2025; Kim et al., 2011; Purnick & Weiss, 2009). This modular approach facilitates the design and assembly of genetic pathways, providing a framework where individual modules or parts can be independently developed, tested, and then integrated into a larger system. The independence of these modules paves the way for creating hierarchical systems with multiple layers of regulation and control, thus enhancing the overall functionality and versatility of synthetic biological systems.

Another key concept in synthetic biology is **abstraction** (Fig. 1c), which involves organising biological functions across hierarchical levels of complexity (Blount et al., 2012). The establishment of these hierarchies is essential as they allow researchers to focus on a single level of complexity, from chemistry to ecology (Henry & Bergelson, 2025; Sebé-Pedrós et al., 2025; Wiegand et al., 2025), without needing to consider details of other levels. At the same time, they facilitate the principled exchange of information across these levels, ensuring coherence and integration within the system. By employing abstraction, parts and devices within engineered biological systems can be redesigned and built to be simpler to model and easier to combine. This approach streamlines the design process, making it more efficient to develop and implement complex genetic circuits and pathways, ultimately enhancing the scalability and robustness of synthetic biological applications.

One more important idea brought from the engineering field is **orthogonality** (Fig. 1d), that allows the modular construction and predictable behaviour of genetic

circuits and pathways (Blount et al., 2012; Costello & Badran, 2021; Edelstein et al., 2025; Wang et al., 2011). Orthogonal components function within interacting the host cell's endogenous systems. This enhances the reliability and robustness of engineered biological systems by minimising crosstalk and interference between components, making systems less susceptible to unintended interactions and more resilient to environmental variations or genetic perturbations, ensuring reproducibility. By assembling orthogonal modules, genetic circuits assembled with standardised parts can be used with confidence, knowing that each component will function independently and contribute to the overall system as intended. Orthogonal components can be stacked to build hierarchical systems with multiple layers of regulation and control, leading to enhanced functionality and versatility.

By applying these concepts, it is possible to develop biological logic gates: standardised genetically engineered modules that integrate one biological input, like transcriptional activation, or protein signalling, and produce a defined output like fluorescence, or gene expression. Biological logic gates are based on Boolean logical rules (AND, OR, NOT…) and can be used to construct more complex genetic or protein-based circuits, thus establishing the foundations of synthetic biology (Brophy & Voigt, 2014; Gander et al., 2017; Gharios et al., 2025; Wang et al., 2011; Wang et al., 2025).

Initiatives such as the International Genetically Engineered Machine (iGEM) competition (https://competition.igem.org/) and the BioBrick Foundation (Røkke et al., 2014) have played a significant role in promoting the adoption of these concepts within the synthetic biology community. These efforts have led to the development of repositories, such as the Registry of Standard Biological Parts (https://technology.igem.org/), Addgene (https://www.addgene.org/) or SynBioHub (https://synbiohub.org/), where researchers can access and share standardised DNA parts for use in their projects.

Finally, by integrating these engineering concepts, we arrive at a crucial methodology in synthetic biology: the **Design-Build-Test-Learn (DBTL) cycle** (Fig. 2). The DBTL cycle is a fundamental framework in synthetic biology that allows researchers to methodically engineer biological systems while iteratively refining their designs

based on experimental feedback (Carbonell et al., 2018; Chao et al., 2017; Nielsen & Keasling, 2016). The cycle consists of four stages or phases: In the **Design phase**, the desired behaviour of the biological system is defined and the genetic modules and genetic modifications necessary to achieve the output are devised *in silico*. Computational tools, such as bioinformatics algorithms and modelling software, aid in the design process by predicting the behaviour of genetic modifications, networks and pathways. Once designed, the required DNA constructs are synthesised or assembled during the **Build phase**. Using standardised parts and modules and the required genetic modifications, the previously designed strain is constructed in a standardised fashion. Once the strain is assembled it is ready to be assessed. In the **Test phase**, the engineered biological system is experimentally tested to evaluate its performance. The easiest way to do this involves measuring the amount, yield or productivity of the desired product, but other parameters should be evaluated in order to obtain a full picture of what is happening in the biological system. Analysing metabolic fluxes or quantifying gene expression are key techniques in the test phase too. Finally, during the **Learn phase**, the performance of the engineered system is studied based on the experimental data collected, and possible improvements are identified. This process can then be iterated cyclically, leading to the refinement and optimisation of the biological system in what is known as the DBTL cycle.

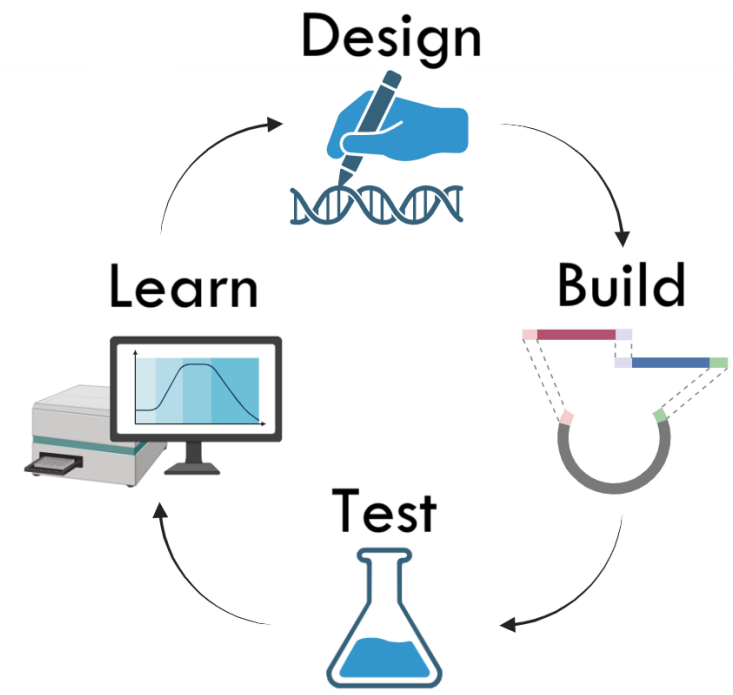


Fig. 2. Schematic representation of the DBTL cycle in synthetic biology. The iterative framework integrates rational design of biological systems, building of such synthetic systems, experimental testing of their performance and learning from their outcome to refine subsequent design rounds, thereby accelerating strain development and biotechnological innovation.

However, for the correct implementation of the DBTL workflow to develop industrial strains, it is crucial to set the right output attributes. To complement the DBTL workflow, a systematic approach widely adopted in biopharmaceutical and biotechnological industries is the concept of Quality by Design (QbD). This concept emphasises the design and development of processes and products with a deep understanding of critical quality attributes (CQAs) and their relationship to critical process parameters (CPPs) (Mu'azzam et al., 2024). Within the context of synthetic biology, QbD ensures the engineered systems and workflows are more robust, predictable, and aligned with the desired performance outcomes. By identifying key variables and their influence early in the design phase, QbD can reduce trial-and-error experimentation, enhance reproducibility, and accelerate the path to optimisation.

Integrating QbD principles into the DBTL cycle allows researchers to incorporate a risk-based framework, where potential sources of variability are anticipated and

mitigated during the design and build phases. This synergy between QbD and DBTL not only enhances the reliability of engineered systems but also aligns with regulatory expectations for scalable and sustainable biotechnological solutions. Together, these methodologies set the scene for more efficient, high-quality innovations in synthetic biology, bridging the gap between research and industrial application.

Cornerstone technologies

The rise of synthetic biology is grounded on the recent expansion of powerful innovations that allow the accurate development of biological systems in a high-throughput manner. The rapid evolution of sequencing technologies, DNA synthesis and manipulation methods as well as bioinformatics and, more recently, artificial intelligence (AI) combined with the structured DBTL workflow and the adoption of engineering concepts, are driving the rapid growth of this field (Cubillos-Ruiz et al., 2021; Lee et al., 2023; Manduva, 2023; Tan et al., 2021; Tang et al., 2021).

One of the first developed technologies that paved the way for synthetic biology was short-read sequencing and particularly Illumina sequencing (Bentley et al., 2008; Margulies et al., 2005). This sequencing method revolutionised genomics by enabling high-throughput, cost-effective DNA sequencing. The substantial reduction in sequencing costs (Fig. 3) and the increase in speed and accuracy in comparison to previous approaches proved to be critical for large-scale genomic studies, allowing researchers to mine the wide diversity of genomes in search of enzymatic activities, alternative chemicals and pathways, opening the road for many biological alternatives that can substitute non-sustainable processes (Caporaso et al., 2012; Carlson, 2025; Fang et al., 2012; Sessou et al., 2019; Shin et al., 2016). Nevertheless, short-read sequencing harbours weaknesses connected to its short-read length: repetitive regions can complicate assembly and structural variant detection, and library construction, usually based on PCR amplification, can introduce mistakes due to secondary DNA structure or high GC content. As sequencing technologies evolved, long-read sequencing methods, such as those developed by Pacific Biosciences (PacBio) (Eid et al., 2009; Fang et al., 2012) and Oxford Nanopore Technologies (ONT) (Akeson et al., 1999; Clarke et al., 2009; Loman et al., 2015), emerged, offering

significant advantages over short-read methods. These long-read technologies produce extensive sequence reads of up to thousands of base pairs, which are crucial for resolving, by themselves or in combination with short-read technologies, complex genomic regions, structural variants, and full-length RNA transcripts. By enhancing our understanding of genomic architecture and facilitating the accurate assembly of synthetic genomes, long-read sequencing has further propelled advancements in synthetic biology and biorefineries (Amarasinghe et al., 2020; Balarezo-Cisneros et al., 2023; Liaquat et al., 2021; Shumate et al., 2022; van der Loos et al., 2021; Wang et al., 2021). Moreover, ONT is now being used for protein sequencing by exploiting ionic current changes as peptides pass through engineered pores (Yusko et al., 2017). This approach allows single-molecule, label-free analysis with the potential to detect amino acid sequences and post-translational modifications in real time (Brinkerhoff et al., 2021; Restrepo-Pérez et al., 2019). Because it avoids bulk degradation and can generate multiple reads of the same molecule, nanopore-based protein sequencing offers a powerful and scalable alternative to mass spectrometry, with particular promise for studying proteoform diversity and enabling portable, high-throughput synthetic biology applications (Wei et al., 2023).

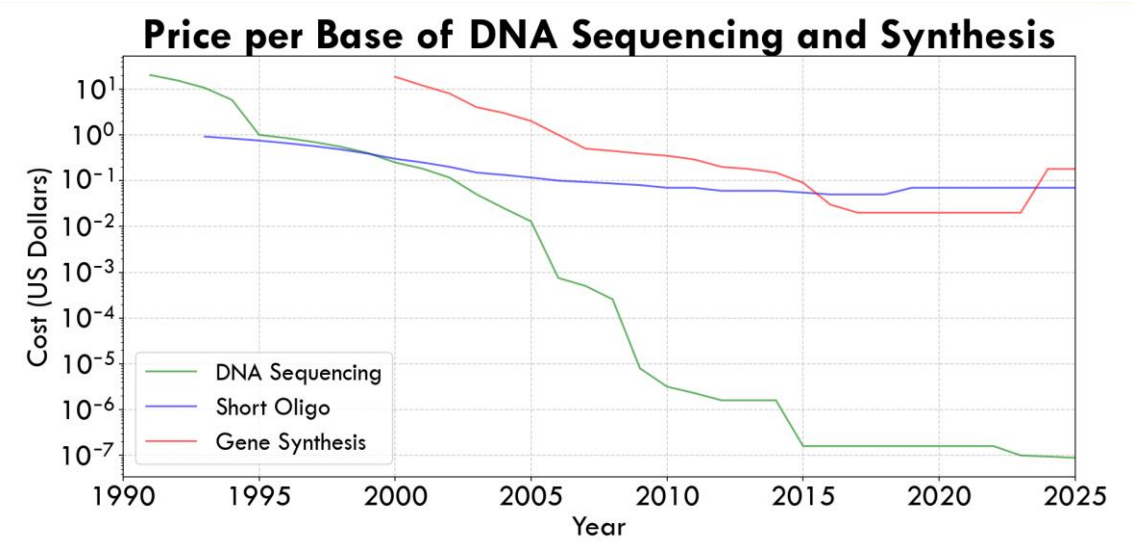


Fig. 3. Historical trends in the cost per base of DNA sequencing (green), short oligonucleotide synthesis (blue), and gene synthesis (red). Adapted from (Carlson, 2025).

Following advances on sequencing, chemical synthesis of DNA has always been recognised as crucial as DNA sequencing. The development of synthetic biology requires the ability of generating long DNA strands containing genes or even full chromosomes. However, despite advances, DNA synthesis remains complex and costly, especially for long sequences (Fig. 3). Currently, synthesis of DNA relies primarily on chemical methods, mainly phosphoramidite synthesis (Beaucage & Caruthers, 1981). This is most efficient for short oligonucleotides, typically up to about <200 nucleotides (bp) in length. As the extent of the DNA strand increases, the efficiency of the synthesis process decreases significantly due to the cumulative probability of errors or incomplete reactions during each cycle of synthesis. Moreover, these truncated sequences that appear during the synthesis step can be difficult to remove completely, affecting the overall purity and yield of the final product. For DNA fragments longer than, approximately 200 base pairs, enzymatic and hybrid methods that join smaller chemically synthesised sequences are the most frequent approach, allowing the concatenation of many segments of hundreds of nucleotides (Hoose et al., 2023; Kosuri & Church, 2014) and this is the current technology in the market (https://www.twistbioscience.com/), with sizes up to 5 kbp. However, new fully enzymatic approaches are earning traction (Palluk et al., 2018), promising similar prizes with higher precision and increased strand length up to 50 kbp (https://ansabio.com/). Despite of the difficulty of constructing long DNA molecules, different synthetic-genome projects appeared at the beginning of 21st century, starting with the landmark complete chemical synthesis of *Mycoplasma genitalium*'s genome and posterior genome transplant for the creation of the first laboratory synthesised bacterium (Fraser et al., 1995; Gibson et al., 2008, 2010; Glass et al., 2006; Lartigue et al., 2007). This research opened the field for *à la carte* design of microorganisms (Angermayr et al., 2015; Galanie et al., 2015; South et al., 2019). Other similar projects include the recent *Escherichia coli* syn61 (Fredens et al., 2019; Robertson et al., 2021; Wang et al., 2016) and *Saccharomyces cerevisiae* 2.0 (Sc2.0) projects (Annaluru et al., 2014; Dymond et al., 2011; Mitchell et al., 2017; Shen et al., 2017; Xie et al., 2017) or more recently even larger and more complex DNA molecules as human chromosomes (Gambogi et al., 2024). These projects collectively aim to advance our understanding of minimal genomes and the fundamental principles of

life by constructing highly engineered or even fully synthetic organisms that could be designed for specific sustainable processes (Claassens et al., 2020; Lee et al., 2019; Ostrov et al., 2016). Key innovations include tools like genetic recoding (Mandell et al., 2015; Robertson et al., 2021), SCRaMbLE (synthetic chromosome rearrangement and modification by loxP-mediated evolution), and the strategic removal of non-essential elements to streamline the genome that could have direct implications in the development of repurposed microorganisms (Dai et al., 2020; Zhang et al., 2021).

Therefore, although phosphoramidite DNA synthesis and variations have many limitations like the cost of reagents, fidelity, the need for sophisticated equipment to control the synthesis environment or the maximum length of 5,000 bp for genes fragments, they still are the preferred method for large-scale synthesis of oligonucleotides but enzymatic methods promise to overcome throughput and length bottlenecks in a field pivoting from chemical synthesis. Continued innovation in this area is essential to make DNA synthesis more accessible and economical, thereby accelerating synthetic biology research and applications.

Together with the ability of writing synthetic DNA sequences, another important goal achieved is the hoarding of a set of tools that allows assembling of large, synthetic DNA constructs. In particular two methods for the in vitro combining of DNA parts or modules have been especially successful: Golden Gate Assembly and Gibson Assembly. Golden Gate Assembly (Engler et al., 2008) leverages type IIS restriction enzymes to facilitate the scarless assembly of multiple DNA fragments. These enzymes cut outside their recognition sites, enabling the precise and efficient construction of complex DNA constructs. This method is widely used in cloning and synthetic biology for its high efficiency and specificity. Its standardisation and modularisation gave rise to one of the most versatile tools for assembling large DNA molecules (Weber et al., 2011) setting a cornerstone for the development of many molecular toolkits for combinatorial assembly of DNA parts in diverse organisms (Crozet et al., 2018; Engler et al., 2014; Fonseca et al., 2019; Lee et al., 2015; Mózsik et al., 2021). Some of its downsides are the need of restriction enzymes sites, the reliance on specific DNA sequences (overhangs) that can affect the output of standard modules and the requirement for entry level parts to be standardised by

removing possible internal restriction enzyme sites. Another successful DNA assembly protocol is Gibson Assembly (Gibson et al., 2009). This method joins DNA fragments by creating overlapping ends that are annealed and repaired by polymerase extension and ligation, circumventing the need for restriction sites, standardisation of parts and offering greater flexibility in construct design. Gibson Assembly has also been standardised by specifying the overlapping ends, offering the possibility of high throughput assembly although with some limitations at sub-gene level (Casini et al., 2014). Other methods like BioBricks™, Gateway cloning or Transformation-Associated Recombination (TAR) represent alternative methods that can be applied in more specific contexts (Ellis et al., 2011; Knight, 2003; Kouprina & Larionov, 2006; Shetty et al., 2008; Walhout et al., 2000). These assembly methods leveraged the capacity of many organisms to combine different DNA parts and apply the DBTL cycle.

Different methods have been developed in order to edit DNA with precision both *in vivo* and *in vitro*, but no method was as successful as the CRISPR-Cas9 technology (Jinek et al., 2012; Mali et al., 2013; Mojica et al., 2005). CRISPR-Cas9 has revolutionised genetic engineering by enabling precise, efficient genome editing and gene expression control. The possibilities are almost endless, allowing the insertion or deletion of full genes down to editing of single nucleotides (Li et al., 2016; Ma et al., 2017). Moreover, it is applied *in vivo*, can be used in a great variety of organisms, and it allows multiplexing of different targets within the cell host (Abdallah et al., 2022; Kabadi et al., 2014; Li et al., 2018; Ousterout et al., 2015). The system uses a guide RNA (gRNA) to direct the Cas9 nuclease to specific DNA sequences, creating double-strand breaks that can be repaired by DNA recombination to introduce mutations or insertions. This technology has become indispensable in synthetic biology for its ease of use and precision, allowing the easy cloning of DNA assemblies constructed with previously described technologies (Chen et al., 2017; Cobb et al., 2015; Hopes et al., 2016; Lowder et al., 2015; Port & Bullock, 2016). Soon after its discovery and application, CRISPR-Cas9 was repurposed for other goals, like the silencing or activation of genes (CRISPRai) (Qi et al., 2013; Shaw et al., 2022), single-base editing without the need of introducing a double-strand DNA nicks (Base-

editing) (Komor et al., 2016) or *in situ* rewriting of the DNA molecule by reverse transcriptase action and ligation (Prime-editing) (Anzalone et al., 2019; Yarnall et al., 2022). Some CRISPR variations allow also for RNA modification without the need to edit the DNA (East-Seletsky et al., 2016; Wang et al., 2025). These methods offer higher precision and fewer off-target effects compared to traditional CRISPR-Cas9 by avoiding double-strand DNA breaks and subsequent recombination, expanding the possibilities for synthetic biology applications (Aznar-Moreno & Durrett, 2017; Datsomor et al., 2019; Galanie et al., 2015; Kuivanen et al., 2016; Yuan et al., 2024).

There are alternative methods based on other enzymes rather than Cas-like that allow insertion, excision and inversion. For instance, new approaches that use RNA-guided recombinases are already available for *E. coli* and have the potential to be brought to other microorganisms (Durrant et al., 2024; Hiraizumi et al., 2024). This system has the advantage of relying on a bi-specific bridge RNA that complements both the donor and the recipient DNA strands. Furthermore, it does not lean on cell-repair mechanisms, leading to safer genome edits and, even more, solving one of the main challenges of CRISPR when trying to apply it to non-model organisms: its dependence on host's homologous recombination machinery (Yan & Fong, 2017).

The engineering of biological systems is not limited to the organism level. While early synthetic biology efforts focused on monocultures, there is growing interest in engineering microbial communities to perform complex tasks collectively (Brenner et al., 2008; McCarty & Ledesma-Amaro, 2019; Wang et al., 2024). These consortia-based approaches can distribute metabolic burdens across different strains or species, exploit natural ecological interactions, and achieve higher resilience under dynamic conditions. Engineered microbial consortia have shown promise in areas such as waste degradation, carbon capture, soil health improvement, and probiotic development (Chen et al., 2024; Mousa et al., 2024; Ruan et al., 2024; Wei et al., 2024; Zhang et al., 2021). Advances in synthetic communication modules, orthogonal quorum sensing systems, and ecological organisation strategies allow the design of interdependent microbial systems with defined roles and behaviours (Henry & Bergelson, 2025; Wu et al., 2022). Additionally, tools for metagenomic analysis and single-cell resolution tracking are enabling better characterisation and control of

community dynamics (Jia et al., 2024; Kim et al., 2024). As we move toward more realistic and scalable biological designs, the engineering of microbial consortia is poised to play a central role in synthetic biology applications in health, agriculture, and the environment.

So far, technologies that modify living organisms have been discussed. Notwithstanding, a rapidly emerging branch in synthetic biology is the development and application of cell-free systems. These platforms operate outside living cells, using purified cellular machinery to express genetic circuits or biosynthetic pathways *in vitro* (Garenne et al., 2021; Hodgman & Jewett, 2012; Rice et al., 2025). Cell-free systems offer several advantages, including rapid prototyping, decoupling of circuit design from cellular metabolism, improved scalability and simplified biosafety. Additionally, cell-free systems are being employed to produce biomolecules or biosensors operating under controlled conditions, in environments that are toxic for living organisms (Grubbe et al., 2020; Kay & Jewett, 2020; Wang et al., 2024). As synthetic biology continues to expand, cell-free systems are proving to be powerful tools for both fundamental research and practical applications in diagnostics, bioremediation, and biomanufacturing.

Building on the discussion of pivotal synthetic biology tools, bioinformatics has proven indispensable for the analysis, interpretation, and visualisation of the vast and complex biological data these technologies generate. Genome assembly software, sequence alignment programs or databases are few examples of crucial tools for managing the huge amount of data. The development of custom programming scripts, data analysis pipelines and online repositories has been essential for processing and interpreting these datasets efficiently. These computational frameworks permit workflow automation, perform complex data analyses, and integrate diverse types of biological information.

For example, developing computational models for synthetic organisms has become an important step in synthetic biology since these models provide a framework to predict and optimise how genetic modifications, metabolic fluxes, and environmental interactions affect the host's overall metabolism and growth, thereby guiding experimental design and reducing the trial-and-error approach (Lopatkin & Collins,

2020). Ultimately, computational modelling accelerates the development of sustainable biotechnology by making the engineering of organisms more efficient, predictable, and scalable (García-Jiménez et al., 2021; Mandakovic et al., 2020; Nielsen & Keasling, 2016; Wagner et al., 2021).

In recent years in particular, with the development of AI and Large Language Models (LLMs) the field of bioinformatics is under a new revolution (Fig. 4). AI offers unprecedented capabilities in data analysis and predictive modelling, being able to find associations barely perceptible to the human eye (Korot et al., 2021). One of the most significant advancements in this domain was AlphaFold, an AI tool developed by DeepMind capable of predicting 3D protein structures from their amino acid sequence (Jumper et al., 2021; Senior et al., 2020). AlphaFold has achieved a remarkable breakthrough in predicting protein structures from amino acid sequences with unparalleled accuracy. This capability allows for the design of novel proteins and enzymes with specific, desired functions. The release of AlphaFold in 2018 represented a turning point in the field of synthetic biology (Fig. 4). AlphaFold's success is, however, just an example of the immense potential of AI in advancing our understanding of biological systems and the integration of AI in bioinformatics extends beyond protein structure prediction. Other recent examples include CRISPR-GPT, a generative transformer-based model that guides researchers in CRISPR related tasks such as selecting Cas variants, designing guide gRNAs, choosing delivery methods, drafting protocols, and interpreting results (Qu et al., 2025) or the development of a pipeline for automatic identification of secretion signals in both prokaryotic and eukaryotic organisms (Teufel et al., 2022). Other machine learning algorithms are increasingly used to identify patterns in large datasets (Chaudhary et al., 2018; Jung et al., 2025), develop enhanced proteins (Huang et al., 2021; Yang et al., 2019), or design genetic circuits (Nielsen et al., 2016; Zou et al., 2024). In addition, the development of AI and machine learning is boosting the automation of research (Carbonell et al., 2019; Stephenson et al., 2023). By integrating robotics, liquid-handling platforms, and automated data collection with computational modelling, automation reduces experimental variability, accelerates strain engineering, and allows the exploration of vast genetic design spaces that would be impractical

manually (Gong et al., 2026). These advances have the potential to bring synthetic biology to a new era of discoveries, enabling the systematic and high-throughput implementation of the DBTL cycle. As AI continues to evolve, its integration with bioinformatics will likely lead to even more sophisticated tools and methodologies, further pushing the boundaries of what is possible in synthetic biology.

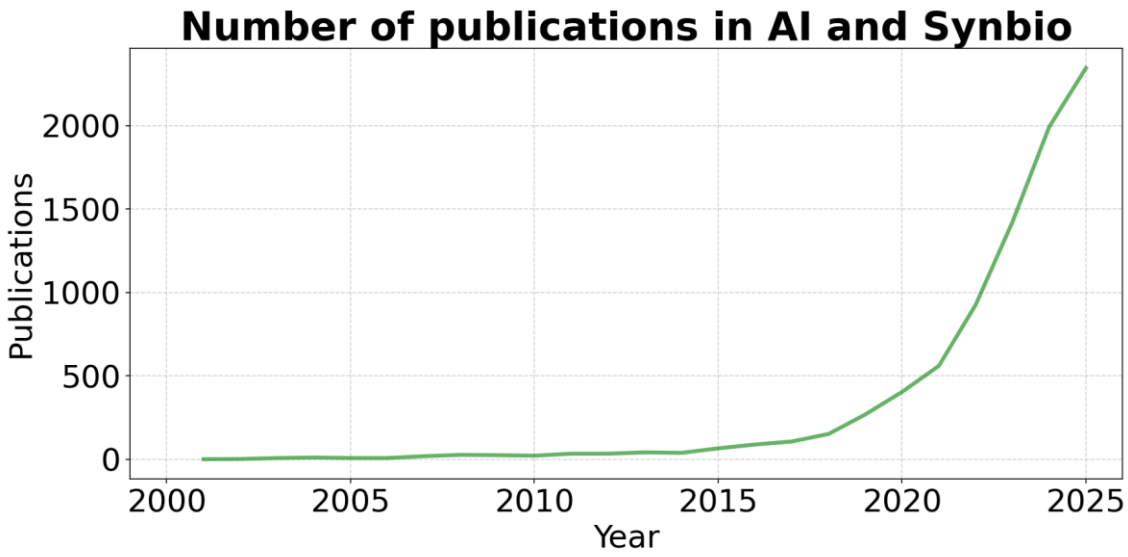


Fig. 4. Evolution of the number of publications retrieved from Scopus (query terms: "synthetic biology" OR related synonyms AND "Artificial Intelligence" OR related synonyms; publication years 2000–2025). Data were exported in September 2025.

Sustainable applications of Synthetic Biology

By abstracting sustainability problems and applying solutions found in nature, synthetic biology is contributing to circularity goals. At present, many efforts try to get inspiration from natural solutions or even mimic them to address problems in a sustainable way. This biomimicry is recurrent in synthetic biology, and many times shapes the direction research takes (Himel et al., 2023; Liu et al., 2022; Waeterschoot et al., 2024). For example, many microbial hosts have been repurposed for sustainable bioproduction of building blocks and biofuels for the industry by integrating heterologous pathways found in other organisms (Antonovsky et al., 2016; Fu et al., 2017; Gutiérrez et al., 2024; Helm et al., 2024; Keasling et al., 2021; Lam & Michael, 2022; Larroude et al., 2018; Liew et al., 2022; Peng et al., 2023;

Walker et al., 2024). These strains implement different renewable approaches like fixing carbon dioxide, using waste feedstocks or providing alternatives to the exploitation of natural resources. Similarly, the development of microbial platforms for the production of high-value chemicals for the pharmaceutical industry, such as artemisinin or cannabinoids, highlights the versatility of synthetic biology in biopharmaceutical manufacturing (Gülck et al., 2020; Kobayashi et al., 2021; Luo et al., 2019; Paddon et al., 2013; Ro et al., 2006).

Synthetic biology approaches are also being applied to agricultural biotechnology. One of the most widespread technologies to increase crop yield is to modify plants so that they become resistant to different plagues and diseases. These strategies are not new, as the earliest started with Bt-corn, a genetically modified corn that expresses *Bacillus thuringiensis* endotoxin which targets and kill certain pest insects, like the European corn borer *Ostrinia nubilalis*, but they sophisticated with time (Huang et al., 2022; Mateos Fernández et al., 2022; Sundararajan et al., 2025; Vaeck et al., 1987; Zhu et al., 2024). Other well-established alternatives to increase crop yields aim at modifying plants to exhibit improved traits such as heat, cold or salt resistances and tolerances, higher nutrient content or enhanced efficiency (Li et al., 2021; Rönspies et al., 2021; Wen et al., 2021; Yang & Reyna-Llorens, 2023; Zhang et al., 2021). Nonetheless, newer approaches are focusing on altering the root-associated microbiome in an effort to circumvent GMO legislation (Jansson et al., 2023; Ke et al., 2021; Shulse et al., 2019; Wen et al., 2021). However, the biggest food production revolution promoted by synthetic biology until now could be the development of novel ways of producing proteins, in particular animal-free meat, and dairy and other products in microbial systems (Biermann et al., 2025; Järviö et al., 2021; Post, 2012; Walker et al., 2024; Xue et al., 2024). Two aspects that have recently caught the attention are precision fermentation (Bae et al., 2025; Hettinga & Bijl, 2022; Mirsalami & Mirsalami, 2025; Teng et al., 2021), which often refers to the heterologous production of animal-related proteins in microbes, and cultivated meat (Lambert et al., 2024; Lei et al., 2023; Reiss et al., 2021), referring to laboratory-grown animal cells that develop into food products (Ancelin et al., 2025). These applications, although far from providing the market with sufficient alternatives to

traditionally grown foods is accelerating and many companies already developed market substitutes for dairy, eggs, meat, palm oil, seafood and other ingredients (https://gfieurope.org/industry).

Another important field where synthetic biology has been applied and keeps growing is the heterologous production of proteins for therapeutical purposes. This application started first with the production of insulin in *E. coli* (Goeddel et al., 1979) and moved then to a variety of target proteins and hosts, like other bacteria, yeast and fungi or plants; and proteins such as growth hormones, clotting factors, or monoclonal antibodies (Balasubramanian et al., 2021; Cutolo et al., 2022; Ghag et al., 2021; Liu et al., 2023; Walsh & Walsh, 2022). The greatest challenges in this field have traditionally been post-translation modifications and processing of the proteins, such as glycosylation , folding or secretion (Bachhav et al., 2023; Listov et al., 2024; Zhao et al., 2024) although the advent of synthetic biology and the emergence of AI has opened new avenues for research, in particular in the design of *de novo* proteins with desired functions (Kortemme, 2024; Vázquez Torres et al., 2025) but also wet-lab strategies that focus on complex problems addressed by synthetic biology, like glycosylation engineering (Liu et al., 2018) or folding and secretion (de Ruijter et al., 2016; Xu et al., 2025).

Synthetic biology also offers new solutions for bioremediation, pollution management, and waste treatment by enabling the precise engineering of biological systems that detect, degrade, or sequester environmental contaminants such as heavy metals, hydrocarbons, pesticides, carbon dioxide, or plastics into non-toxic compounds (Brandenberg et al., 2022; Cowan et al., 2025; Lorenzo et al., 2018; Ren et al., 2025; Soon et al., 2025) or even high added value molecules (Johnson et al., 2025). For instance, engineered biosensors can be programmed to detect trace levels of contaminants with high specificity (Dong et al., 2017; Garabello et al., 2024; Kumar et al., 2017), while microbial consortia can be optimised for cooperative degradation of complex waste streams (Bao et al., 2023). These capabilities not only improve current waste management strategies but also pave the way for regenerative approaches that convert waste into valuable bioproducts.

Challenges

Despite promising advances, synthetic biology faces significant challenges, the greatest being its final goal: reaching the level of complexity of a living cell. Within a single cell, thousands of biochemical processes occur simultaneously, coordinated by regulatory networks that have been refined over hundreds of millions of years. These networks are deeply interconnected and have evolved intricate layers of feedback, robustness, and adaptability that are difficult to engineer from scratch. To replicate the complexity achieved by natural evolution and benefit from it, synthetic biology needs to fully understand and reconstruct the natural complexity for human purposes (Bervoets & Charlier, 2019; O'Malley et al., 2008). Notwithstanding, great advantages have been done by abstracting the problems together with the aid of AI (Adamala et al., 2024; Lu et al., 2024).

Another important challenge is to scale the production of laboratory achievements to commercialisation level (Kampers et al., 2021). Bioprocesses require cost-effective and high-yield systems, especially for bioprocesses aiming at low added-value products like bioethanol, cultured meat or bulk chemicals that need to compete with their traditional counterparts (Danchin, 2012; Helm et al., 2024; Humbird, 2021; Liew et al., 2022; Rasul et al., 2022). For instance, the adoption of non-traditional organisms as production hosts in order to address these requirements, although offering technical advantages, also requires other solutions for different problems that arise during their scale up (Al-Hammadi et al., 2025; Coughlan et al., 2022; Doğan et al., 2023; Hong et al., 2022; Poppeliers et al., 2025). Many times, when focus is on the research, these needs that industrial set-ups demand can be neglected. When this research is scaled up, many factors can differ, these differences between laboratory and industrial set-ups become clearer when critical performance parameters cannot be maintained. These factors vary depending on the bioprocess, but the most common reasons identified in different bioprocesses are: the formation of gradients along the bioreactor (substrate, temperature, oxygen…) (Becker et al., 2023; Bylund et al., 1998; Spadiut et al., 2013), the consequently generation of heterogeneous subpopulations that adapt to the microenvironments in the bioreactor (Delvigne & Goffin, 2014; Shabestary et al., 2024), the industrial conditions

that were not considered during the strain construction (Becker et al., 2023; Olsson et al., 2022), and a suboptimal selection of DNA parts due to the changing conditions during the scale-up path (Garrigós-Martínez et al., 2021; Georgiadis et al., 2023; Kittler et al., 2025; Pan et al., 2023). The so-called “scale-up effect” is due to a combination of these factors affecting the performance of laboratory developed processes. To tackle this problem, some strategies focus on the development of robust chassis strains that withstand the operational variation in industrial set-ups (Gong et al., 2017; Olsson et al., 2022). However, identification of crucial steps during development of bioprocess is a key factor for their success. QbD offers a structured framework to improve the robustness and scalability of synthetic biology-based systems, emphasising a deep understanding of the process and product by identifying CQAs and CPPs from early stages, helping to ensure that variations during fermentation scale-up do not compromise performance.

Two other challenges synthetic biology faces are not scientific, but socioeconomic. Consumer acceptance plays a critical role in the development of synthetic biology solutions, with public perception influenced by concerns over safety, naturalness, and cultural familiarity with lab-made goods. Public perception profoundly shapes market adoption, investment, and policy decisions, especially in applications that involve food, health, or environmental release, often due to concerns over unnaturalness, safety, or ethical implications (Bearth et al., 2022; GFI, 2025; Hocquette et al., 2025; Lanz et al., 2024). In contrast, applications in industrial biocatalysis or closed-loop bioremediation systems, tend to face less public scrutiny (Olynk Widmar et al., 2017). Transparent communication, strong safety data, and engagement with ethical, cultural, and environmental values will be essential to build trust (Beghin & Gustafson, 2021; Koralesky et al., 2023; Paudel et al., 2023).

Regulations affecting synthetic biology can be complex and vary widely between countries and regions, creating uncertainty for innovators and developers and slowing the global adoption and commercialisation of new biotechnologies (Turnbull et al., 2021; Wesseler et al., 2023; Wong & Chan, 2016). Regulatory policies are primarily driven by health, safety and environmental concerns, with particular attention given to GMOs and their potential ecological impacts. Concerns raised by

regulatory bodies (e.g. European Food Safety Authority, Food and Drug Administration…) often include horizontal gene transfer, persistence in the environment, and unintended consequences of engineered traits (Frankel & Chapman, 2000; Hayashi et al., 2024; More et al., 2022; Mullins et al., 2024; Son et al., 2021). However, recent advances in synthetic biology including the design of streamlined genomes (Leprince et al., 2012), the development of kill switches and auxotrophies (Asin-Garcia et al., 2022; Rottinghaus et al., 2022), and the implementation of tightly controlled gene circuits (Asensio-Calavia et al., 2024), offer new opportunities to address these biosafety concerns, supporting the creation of evidence-based, adaptive regulatory frameworks that balance safety with technological progress.

Conclusion

Synthetic biology has emerged as a transformative discipline by integrating engineering principles into biological design. Enabled by rapid advances in DNA synthesis, sequencing technologies, and molecular editing tools like CRISPR, synthetic biology is reshaping industrial biotechnology. Among its most promising applications is the development of sustainable biorefineries. However, scaling up laboratory innovations, navigating complex regulatory frameworks, and achieving consumer acceptance remain key challenges. Addressing these will require continued emphasis on rational design through frameworks like QbD, as well as stronger interdisciplinary collaboration. As synthetic biology matures, its ability to contribute to circular and sustainable bioeconomy will depend not only on technical breakthroughs, but also on responsible implementation and public engagement.

References

Abdallah, N. A., Elsharawy, H., Abulela, H. A., Thilmony, R., Abdelhadi, A. A., & Elarabi, N. I. (2022). Multiplex CRISPR/Cas9-mediated genome editing to address drought tolerance in wheat. *GM Crops & Food*. https://doi.org/10.1080/21645698.2022.2120313

Adamala, K. P., Dogterom, M., Elani, Y., Schwille, P., Takinoue, M., & Tang, T. Y. D. (2024). Present and future of synthetic cell development. *Nature Reviews Molecular Cell Biology*, *25*(3), 162–167. https://doi.org/10.1038/S41580-023-00686-9

Akeson, M., Branton, D., Kasianowicz, J. J., Brandin, E., & Deamer, D. W. (1999). Microsecond Time-Scale Discrimination Among Polycytidylic Acid, Polyadenylic Acid, and Polyuridylic Acid as Homopolymers or as Segments Within Single RNA Molecules. *Biophysical Journal*, *77*, 3227–3233. https://doi.org/10.1016/S0006-3495(99)77153-5

Al-Hammadi, M., Anadol, G., Martín-García, F. J., Moreno-García, J., Keskin Gündoğdu, T., & Güngörmüşler, M. (2025). Scaling bioethanol for the future: the commercialization potential of extremophiles and non-conventional microorganisms. *Frontiers in Energy Research*, *13*, 1565273. https://doi.org/10.3389/FENRG.2025.1565273

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biology 2020 21:1*, *21*(1), 1–16. https://doi.org/10.1186/S13059-020-1935-5

Ancelin, M., Martins dos Santos, V. A. P., Morrissey, J. P., O'Donohue, M. J., Penttilä, M., & Philp, J. C. (2025). Addressing semantic ambiguity in biotechnology: Proposals from the European research infrastructure IBISBA. *New Biotechnology*, *88*, 83–88. https://doi.org/10.1016/J.NBT.2025.04.010

Angermayr, S. A., Gorchs Rovira, A., & Hellingwerf, K. J. (2015). Metabolic engineering of cyanobacteria for the synthesis of commodity products. *Trends in Biotechnology*, *33*(6), 352–361. https://doi.org/10.1016/j.tibtech.2015.03.009

Annaluru, N., Muller, H., Mitchell, L. A., Ramalingam, S., Stracquadanio, G., Richardson, S. M., Dymond, J. S., Kuang, Z., Scheifele, L. Z., Cooper, E. M., Cai, Y., Zeller, K., Agmon,

N., Han, J. S., Hadjithomas, M., Tullman, J., Caravelli, K., Cirelli, K., Guo, Z., … Chandrasegaran, S. (2014). Total synthesis of a functional designer eukaryotic chromosome. *Science*, *344*(6179), 55–58. https://doi.org/10.1126/SCIENCE.1249252

Antonovsky, N., Gleizer, S., Noor, E., Zohar, Y., Herz, E., Barenholz, U., Zelcbuch, L., Amram, S., Wides, A., Tepper, N., Davidi, D., Bar-On, Y., Bareia, T., Wernick, D. G., Shani, I., Malitsky, S., Jona, G., Bar-Even, A., & Milo, R. (2016). Sugar Synthesis from CO2 in Escherichia coli. *Cell*, *166*(1), 115–125. https://doi.org/10.1016/J.CELL.2016.05.064

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature 2019 576:7785*, *576*(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4

Asensio-Calavia, A., Ceballos-Munuera, Á., Méndez-Pérez, A., Álvarez, B., & Fernández, L. Á. (2024). A tuneable genetic switch for tight control of tac promoters in Escherichia coli boosts expression of synthetic injectisomes. *Microbial Biotechnology*, *17*(1), e14328. https://doi.org/10.1111/1751-7915.14328

Asin-Garcia, E., Batianis, C., Li, Y., Fawcett, J. D., de Jong, I., & dos Santos, V. A. P. M. (2022). Phosphite synthetic auxotrophy as an effective biocontainment strategy for the industrial chassis Pseudomonas putida. *Microbial Cell Factories 2022 21:1*, *21*(1), 156-. https://doi.org/10.1186/S12934-022-01883-5

Aznar-Moreno, J. A., & Durrett, T. P. (2017). Simultaneous Targeting of Multiple Gene Homeologs to Alter Seed Oil Production in Camelina sativa. *Plant and Cell Physiology*, *58*(7), 1260–1267. https://doi.org/10.1093/PCP/PCX058

Bachhav, B., de Rossi, J., Llanos, C. D., & Segatori, L. (2023). Cell factory engineering: Challenges and opportunities for synthetic biology applications. *Biotechnology and Bioengineering*, *120*(9), 2441–2459. https://doi.org/10.1002/BIT.28365

Bae, H. N., Kim, G. H., & Seo, S. O. (2025). Secretory Production of Plant Heme-Containing Globins by Recombinant Yeast via Precision Fermentation. *Foods*, *14*(8), 1422. https://doi.org/10.3390/FOODS14081422

Balarezo-Cisneros, L. N., Timouma, S., Hanak, A., Currin, A., Valle, F., & Delneri, D. (2023). High quality de novo genome assembly of the non-conventional yeast Kazachstania bulderi describes a potential low pH production host for biorefineries. *Communications Biology 2023 6:1*, *6*(1), 1–14. https://doi.org/10.1038/s42003-023-05285-0

Balasubramanian, S., Chen, J., Wigneswaran, V., Bang-Berthelsen, C. H., & Jensen, P. R. (2021). Droplet-Based Microfluidic High Throughput Screening of Corynebacterium glutamicum for Efficient Heterologous Protein Production and Secretion. *Frontiers in Bioengineering and Biotechnology*, *9*, 668513. https://doi.org/10.3389/FBIOE.2021.668513

Bao, T., Qian, Y., Xin, Y., Collins, J. J., & Lu, T. (2023). Engineering microbial division of labor for plastic upcycling. *Nature Communications*, *14*(1), 1–13. https://doi.org/10.1038/S41467-023-40777-X

Bearth, A., Kaptan, G., & Kessler, S. H. (2022). Genome-edited versus genetically-modified tomatoes: an experiment on people's perceptions and acceptance of food biotechnology in the UK and Switzerland. *Agriculture and Human Values*, *39*(3), 1117–1131. https://doi.org/10.1007/S10460-022-10311-8

Beaucage, S. L., & Caruthers, M. H. (1981). Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters*, *22*(20), 1859–1862. https://doi.org/10.1016/S0040-4039(01)90461-7

Becker, L., Sturm, J., Eiden, F., & Holtmann, D. (2023). Analyzing and understanding the robustness of bioprocesses. *Trends in Biotechnology*, *41*(8), 1013–1026. https://doi.org/10.1016/J.TIBTECH.2023.03.002

Beghin, J. C., & Gustafson, C. R. (2021). Consumer valuation of and attitudes towards novel foods produced with new plant engineering techniques: A review. *Sustainability (Switzerland)*, *13*(20), 11348. https://doi.org/10.3390/SU132011348

Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Hall, K. P., Evers, D. J., Barnes, C. L., Bignell, H. R., Boutell, J. M., Bryant, J., Carter, R. J., Keira Cheetham, R., Cox, A. J., Ellis, D. J., Flatbush, M. R., Gormley, N. A., Humphray,

S. J., … Smith, A. J. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature 2008 456:7218*, *456*(7218), 53–59. https://doi.org/10.1038/nature07517

Bervoets, I., & Charlier, D. (2019). Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. *FEMS Microbiology Reviews*, *43*(3), 304–339. https://doi.org/10.1093/FEMSRE/FUZ001

Biermann, L., Tadele, L. R., Henrique, E., Perino, B., Nicholson, R., Lilge, L., & Hausmann, R. (2025). Recombinant Production of Bovine αS1-Casein in Genome-Reduced Bacillus subtilis Strain IIG-Bs-20-5-1. *Microorganisms 2025, Vol. 13, Page 60*, *13*(1), 60. https://doi.org/10.3390/MICROORGANISMS13010060

Blount, B. A., Weenink, T., Vasylechko, S., & Ellis, T. (2012). Rational Diversification of a Promoter Providing Fine-Tuned Expression and Orthogonal Regulation for Synthetic Biology. *PLOS ONE*, *7*(3), e33279. https://doi.org/10.1371/JOURNAL.PONE.0033279

Brandenberg, O. F., Schubert, O. T., & Kruglyak, L. (2022). Towards synthetic PETtrophy: Engineering Pseudomonas putida for concurrent polyethylene terephthalate (PET) monomer metabolism and PET hydrolase expression. *Microbial Cell Factories*, *21*(1), 1–31. https://doi.org/10.1186/S12934-022-01849-7

Brans, K. I., Jansen, M., Vanoverbeke, J., Tüzün, N., Stoks, R., & De Meester, L. (2017). The heat is on: Genetic adaptation to urbanization mediated by thermal tolerance and body size. *Global Change Biology*, *23*(12), 5218–5227. https://doi.org/10.1111/GCB.13784

Brenner, K., You, L., & Arnold, F. H. (2008). Engineering microbial consortia: a new frontier in synthetic biology. *Trends in Biotechnology*, *26*(9), 483–489. https://doi.org/10.1016/J.TIBTECH.2008.05.004

Brinkerhoff, H., Kang, A. S. W., Liu, J., Aksimentiev, A., & Dekker, C. (2021). Multiple rereads of single proteins at single–amino acid resolution using nanopores. *Science*, *374*(6574), 1509–1513. https://doi.org/10.1126/SCIENCE.ABL4381

Brophy, J. A. N., & Voigt, C. A. (2014). Principles of genetic circuit design. *Nature Methods 2014 11:5*, *11*(5), 508–520. https://doi.org/10.1038/nmeth.2926

Butti, P., Bellusci, F., Brambilla, E., & Branduardi, P. (2025). Genomically integrated cassettes swapping: bringing modularity to the strain level in Saccharomyces cerevisiae. *FEMS Yeast Research*, *25*, 32. https://doi.org/10.1093/FEMSYR/FOAF032

Bylund, F., Collet, E., Enfors, S. O., & Larsson, G. (1998). Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. *Bioprocess Engineering*, *18*(3), 171–180. https://doi.org/10.1007/S004490050427

Canton, B., Labno, A., & Endy, D. (2008). Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology 2008 26:7*, *26*(7), 787–793. https://doi.org/10.1038/nbt1413

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S. M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J. A., Smith, G., & Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, *6*(8), 1621–1624. https://doi.org/10.1038/ISMEJ.2012.8

Carbonell, P., Jervis, A. J., Robinson, C. J., Yan, C., Dunstan, M., Swainston, N., Vinaixa, M., Hollywood, K. A., Currin, A., Rattray, N. J. W., Taylor, S., Spiess, R., Sung, R., Williams, A. R., Fellows, D., Stanford, N. J., Mulherin, P., Le Feuvre, R., Barran, P., … Scrutton, N. S. (2018). An automated Design-Build-Test-Learn pipeline for enhanced microbial production of fine chemicals. *Communications Biology 2018 1:1*, *1*(1), 1–10. https://doi.org/10.1038/s42003-018-0076-9

Carbonell, P., Radivojevic, T., & García Martín, H. (2019). Opportunities at the Intersection of Synthetic Biology, Machine Learning, and Automation. *ACS Synthetic Biology*, *8*(7), 1474–1477. https://doi.org/10.1021/ACSSYNBIO.8B00540

Carlson, R. (2025, May). *Synthesis*. https://www.synthesis.cc/synthesis/category/Carlson+Curves

Casini, A., Macdonald, J. T., Jonghe, J. De, Christodoulou, G., Freemont, P. S., Baldwin, G. S., & Ellis, T. (2014). One-pot DNA construction for synthetic biology: the Modular

Overlap-Directed Assembly with Linkers (MODAL) strategy. *Nucleic Acids Research*, *42*(1), e7–e7. https://doi.org/10.1093/NAR/GKT915

Chao, R., Mishra, S., Si, T., & Zhao, H. (2017). Engineering biological systems using automated biofoundries. *Metabolic Engineering*, *42*, 98–108. https://doi.org/10.1016/J.YMBEN.2017.06.003

Chaudhary, K., Poirion, O. B., Lu, L., & Garmire, L. X. (2018). Deep learning–based multi-omics integration robustly predicts survival in liver cancer. *Clinical Cancer Research*, *24*(6), 1248–1259. https://doi.org/10.1158/1078-0432.CCR-17-0853

Chen, W., Zhang, Y., Yeo, W. S., Bae, T., & Ji, Q. (2017). Rapid and Efficient Genome Editing in Staphylococcus aureus by Using an Engineered CRISPR/Cas9 System. *Journal of the American Chemical Society*, *139*(10), 3790–3795. https://doi.org/10.1021/JACS.6B13317

Chen, X., Gao, M., Wang, L., Qiang, G., Wu, Y., Huang, H., & Kang, G. (2024). A synthetic microbial consortium protects against obesity by regulating vitamin B6 metabolism. *Gut Microbes*, *16*(1). https://doi.org/10.1080/19490976.2024.2304901

Claassens, N. J., Bordanaba-Florit, G., Cotton, C. A. R., De Maria, A., Finger-Bou, M., Friedeheim, L., Giner-Laguarda, N., Munar-Palmer, M., Newell, W., Scarinci, G., Verbunt, J., de Vries, S. T., Yilmaz, S., & Bar-Even, A. (2020). Replacing the Calvin cycle with the reductive glycine pathway in Cupriavidus necator. *Metabolic Engineering*, *62*, 30–41. https://doi.org/10.1016/J.YMBEN.2020.08.004

Clarke, J., Wu, H. C., Jayasinghe, L., Patel, A., Reid, S., & Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology 2009 4:4*, *4*(4), 265–270. https://doi.org/10.1038/nnano.2009.12

Cobb, R. E., Wang, Y., & Zhao, H. (2015). High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synthetic Biology*, *4*(6), 723–728. https://doi.org/10.1021/SB500351F

Costello, A., & Badran, A. H. (2021). Synthetic Biological Circuits within an Orthogonal Central Dogma. *Trends in Biotechnology*, *39*(1), 59–71. https://doi.org/10.1016/J.TIBTECH.2020.05.013

Coughlan, N. E., Walsh, É., Bolger, P., Burnell, G., O'Leary, N., O'Mahoney, M., Paolacci, S., Wall, D., & Jansen, M. A. K. (2022). Duckweed bioreactors: Challenges and opportunities for large-scale indoor cultivation of Lemnaceae. *Journal of Cleaner Production*, *336*, 130285. https://doi.org/10.1016/J.JCLEPRO.2021.130285

Cowan, A. E., Hillers, M., Rainaldi, V., Collas, F., Choudhary, H., Zakaria, B. S., Bieberach, G. G., Carruthers, D. N., Grabovac, M., Gin, J. W., Cawthon, B., Chen, Y., Turumtay, E. A., Baidoo, E. E. K., Petzold, C. J., Feist, A. M., Tejedor-Sanz, S., Kensy, F., Simmons, B. A., … Claassens, N. J. (2025). Fast growth and high-titer bioproduction from renewable formate via metal-dependent formate dehydrogenase in Escherichia coli. *Nature Communications 2025 16:1*, *16*(1), 1–13. https://doi.org/10.1038/s41467-025-61001-y

Crozet, P., Navarro, F. J., Willmund, F., Mehrshahi, P., Bakowski, K., Lauersen, K. J., Pérez-Pérez, M. E., Auroy, P., Gorchs Rovira, A., Sauret-Gueto, S., Niemeyer, J., Spaniol, B., Theis, J., Trösch, R., Westrich, L. D., Vavitsas, K., Baier, T., Hübner, W., De Carpentier, F., … Lemaire, S. D. (2018). Birth of a Photosynthetic Chassis: A MoClo Toolkit Enabling Synthetic Biology in the Microalga Chlamydomonas reinhardtii. *ACS Synthetic Biology*, *7*(9), 2074–2086. https://doi.org/10.1021/ACSSYNBIO.8B00251

Cubillos-Ruiz, A., Guo, T., Sokolovska, A., Miller, P. F., Collins, J. J., Lu, T. K., & Lora, J. M. (2021). Engineering living therapeutics with synthetic biology. *Nature Reviews Drug Discovery 2021 20:12*, *20*(12), 941–960. https://doi.org/10.1038/s41573-021-00285-3

Cutolo, E. A., Mandalà, G., Dall'osto, L., & Bassi, R. (2022). Harnessing the Algal Chloroplast for Heterologous Protein Production. *Microorganisms 2022, Vol. 10, Page 743*, *10*(4), 743. https://doi.org/10.3390/MICROORGANISMS10040743

Dahiya, S., Kumar, A. N., Shanthi Sravan, J., Chatterjee, S., Sarkar, O., & Mohan, S. V. (2018). Food waste biorefinery: Sustainable strategy for circular bioeconomy. *Bioresource Technology*, *248*, 2–12. https://doi.org/10.1016/J.BIORTECH.2017.07.176

Dai, J., Boeke, J. D., Luo, Z., Jiang, S., & Cai, Y. (2020). Sc3.0: Revamping and minimizing the yeast genome. *Genome Biology*, *21*(1), 1–4. https://doi.org/10.1186/S13059-020-02130-Z

Danchin, A. (2012). Scaling up synthetic biology: Do not forget the chassis. *FEBS Letters*, *586*(15), 2129–2137. https://doi.org/10.1016/J.FEBSLET.2011.12.024

Datsomor, A. K., Olsen, R. E., Zic, N., Madaro, A., Bones, A. M., Edvardsen, R. B., Wargelius, A., & Winge, P. (2019). CRISPR/Cas9-mediated editing of Δ5 and Δ6 desaturases impairs Δ8-desaturation and docosahexaenoic acid synthesis in Atlantic salmon (Salmo salar L.). *Scientific Reports 2019 9:1*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-53316-w

de Ruijter, J. C., Koskela, E. V., & Frey, A. D. (2016). Enhancing antibody folding and secretion by tailoring the Saccharomyces cerevisiae endoplasmic reticulum. *Microbial Cell Factories*, *15*(1), 1–18. https://doi.org/10.1186/S12934-016-0488-5

Delvigne, F., & Goffin, P. (2014). Microbial heterogeneity affects bioprocess robustness: Dynamic single-cell analysis contributes to understanding of microbial populations. *Biotechnology Journal*, *9*(1), 61–72. https://doi.org/10.1002/BIOT.201300119

Doğan, B., Yıldız, Z., Aksöz, N., Eninanç, A. B., Korkmaz Kahveci, B. G., & Yamaç, M. (2023). Optimization and reactor-scale production of plant growth regulators by Pleurotus eryngii. *3 Biotech*, *13*(9), 1–14. https://doi.org/10.1007/S13205-023-03744-3

Dong, Z. Z., Yang, C., Vellaisamy, K., Li, G., Leung, C. H., & Ma, D. L. (2017). Construction of a Nano Biosensor for Cyanide Anion Detection and Its Application in Environmental and Biological Systems. *ACS Sensors*, *2*(10), 1517–1522. https://doi.org/10.1021/ACSSENSORS.7B00553

Durrant, M. G., Perry, N. T., Pai, J. J., Jangid, A. R., Athukoralage, J. S., Hiraizumi, M., McSpedon, J. P., Pawluk, A., Nishimasu, H., Konermann, S., & Hsu, P. D. (2024). Bridge RNAs direct programmable recombination of target and donor DNA. *Nature 2024 630:8018*, *630*(8018), 984–993. https://doi.org/10.1038/s41586-024-07552-4

Dymond, J. S., Richardson, S. M., Coombes, C. E., Babatz, T., Muller, H., Annaluru, N., Blake, W. J., Schwerzmann, J. W., Dai, J., Lindstrom, D. L., Boeke, A. C., Gottschling, D. E., Chandrasegaran, S., Bader, J. S., & Boeke, J. D. (2011). Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature 2011 477:7365*, *477*(7365), 471–476. https://doi.org/10.1038/nature10403

East-Seletsky, A., O'Connell, M. R., Knight, S. C., Burstein, D., Cate, J. H. D., Tjian, R., & Doudna, J. A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature 2016 538:7624*, *538*(7624), 270–273. https://doi.org/10.1038/nature19802

Edelstein, H. I., Cosio, A., Ezekiel, M. L., Corcoran, W. K., Morris, A. H., & Leonard, J. N. (2025). Conversion of natural cytokine receptors into orthogonal synthetic biosensors. *Nature Chemical Biology 2025*, 1–12. https://doi.org/10.1038/s41589-025-01986-1

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., DeWinter, A., Dixon, J., … Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, *323*(5910), 133–138. https://doi.org/10.1126/SCIENCE.1162986

Ellis, T., Adie, T., & Baldwin, G. S. (2011). DNA assembly for synthetic biology: from parts to pathways and beyond. *Integrative Biology*, *3*(2), 109–118. https://doi.org/10.1039/C0IB00070A

Elowitz, M. B., & Leibier, S. (2000). A synthetic oscillatory network of transcriptional regulators. *Nature 2000 403:6767*, *403*(6767), 335–338. https://doi.org/10.1038/35002125

Endy, D. (2005). Foundations for engineering biology. *Nature 2005 438:7067*, *438*(7067), 449–453. https://doi.org/10.1038/nature04342

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE*, *3*(11), e3647. https://doi.org/10.1371/JOURNAL.PONE.0003647

Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D. G., Patron, N. J., & Marillonnet, S. (2014). A Golden Gate modular cloning toolbox for plants. *ACS Synthetic Biology*, *3*(11), 839–843. https://doi.org/10.1021/SB4001504

Fang, G., Munera, D., Friedman, D. I., Mandlik, A., Chao, M. C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M. C., Jabado, O. J., Deikus, G., Clark, T. A., Luong, K., Murray, I. A., Davis,

B. M., Keren-Paz, A., Chess, A., Roberts, R. J., Korlach, J., … Schadt, E. E. (2012). Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing. *Nature Biotechnology 2012 30:12*, *30*(12), 1232–1239. https://doi.org/10.1038/nbt.2432

Fang, X., Mou, Y., Huang, Z., Li, Y., Han, L., Zhang, Y., Feng, Y., Chen, Y., Jiang, X., Zhao, W., Sun, X., Xiong, Z., Yang, L., Liu, H., Fan, D., Mao, L., Ren, L., Liu, C., Wang, J., … Du, Y. (2012). The sequence and analysis of a Chinese pig genome. *GigaScience*, *1*(1). https://doi.org/10.1186/2047-217X-1-16

Fonseca, J. P., Bonny, A. R., Kumar, G. R., Ng, A. H., Town, J., Wu, Q. C., Aslankoohi, E., Chen, S. Y., Dods, G., Harrigan, P., Osimiri, L. C., Kistler, A. L., & El-Samad, H. (2019). A Toolkit for Rapid Modular Construction of Biological Circuits in Mammalian Cells. *ACS Synthetic Biology*, *8*(11), 2593–2606. https://doi.org/10.1021/ACSSYNBIO.9B00322

Frankel, M. S., & Chapman, A. R. (2000). *Human Inheritable Genetic Modifications Assessing Scientific, Ethical, Religious, and Policy Issues*. http://www.aaas.org/spp/dspp/sfrl/germline/main.htm

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., … Venter, J. C. (1995). The Minimal Gene Complement of Mycoplasma genitalium. *Science*, *270*(5235), 397–403. https://doi.org/10.1126/SCIENCE.270.5235.397

Fredens, J., Wang, K., de la Torre, D., Funke, L. F. H., Robertson, W. E., Christova, Y., Chia, T., Schmied, W. H., Dunkelmann, D. L., Beránek, V., Uttamapinant, C., Llamazares, A. G., Elliott, T. S., & Chin, J. W. (2019). Total synthesis of Escherichia coli with a recoded genome. *Nature 2019 569:7757*, *569*(7757), 514–518. https://doi.org/10.1038/s41586-019-1192-5

Fu, H., Yang, S. T., Wang, M., Wang, J., & Tang, I. C. (2017). Butyric acid production from lignocellulosic biomass hydrolysates by engineered Clostridium tyrobutyricum overexpressing xylose catabolism genes for glucose and xylose co-utilization.

Bioresource Technology, *234*, 389–396. https://doi.org/10.1016/J.BIORTECH.2017.03.073

Galanie, S., Thodey, K., Trenchard, I. J., Interrante, M. F., & Smolke, C. D. (2015). Complete biosynthesis of opioids in yeast. *Science*, *349*(6252), 1095–1100. https://doi.org/10.1126/SCIENCE.AAC9373

Galdzicki, M., Clancy, K. P., Oberortner, E., Pocock, M., Quinn, J. Y., Rodriguez, C. A., Roehner, N., Wilson, M. L., Adam, L., Anderson, J. C., Bartley, B. A., Beal, J., Chandran, D., Chen, J., Densmore, D., Endy, D., Grünberg, R., Hallinan, J., Hillson, N. J., … Sauro, H. M. (2014). The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology. *Nature Biotechnology 2014 32:6*, *32*(6), 545–550. https://doi.org/10.1038/nbt.2891

Gambogi, C. W., Birchak, G. J., Mer, E., Brown, D. M., Yankson, G., Kixmoeller, K., Gavade, J. N., Espinoza, J. L., Kashyap, P., Dupont, C. L., Logsdon, G. A., Heun, P., Glass, J. I., & Black, B. E. (2024). Efficient formation of single-copy human artificial chromosomes. *Science*, *383*(6689), 1344–1349. https://doi.org/10.1126/SCIENCE.ADJ3566

Gander, M. W., Vrana, J. D., Voje, W. E., Carothers, J. M., & Klavins, E. (2017). Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nature Communications*, *8*(1), 1–11. https://doi.org/10.1038/NCOMMS15459

Garabello, E., Yoon, H., Reid, M. C., & Giometto, A. (2024). Tunable Low-Rate Genomic Recombination with Cre-lox in E. coli: A Versatile Tool for Synthetic Biology and Environmental Sensing. *BioRxiv*, 2024.10.02.616356. https://doi.org/10.1101/2024.10.02.616356

García-Jiménez, B., Torres-Bacete, J., & Nogales, J. (2021). Metabolic modelling approaches for describing and engineering microbial communities. *Computational and Structural Biotechnology Journal*, *19*, 226–246. https://doi.org/10.1016/J.CSBJ.2020.12.003

Garenne, D., Thompson, S., Brisson, A., Khakimzhan, A., & Noireaux, V. (2021). The all-E. coliTXTL toolbox 3.0: new capabilities of a cell-free synthetic biology platform. *Synthetic Biology*, *6*(1). https://doi.org/10.1093/SYNBIO/YSAB017

Garrigós-Martínez, J., Vuoristo, K., Nieto-Taype, M. A., Tähtiharju, J., Uusitalo, J., Tukiainen, P., Schmid, C., Tolstorukov, I., Madden, K., Penttilä, M., Montesinos-Seguí, J. L., Valero, F., Glieder, A., & Garcia-Ortega, X. (2021). Bioprocess performance analysis of novel methanol-independent promoters for recombinant protein production with Pichia pastoris. *Microbial Cell Factories*, *20*(1), 1–12. https://doi.org/10.1186/S12934-021-01564-9

Georgiadis, I., Tsiligkaki, C., Patavou, V., Orfanidou, M., Tsoureki, A., Andreadelli, A., Theodosiou, E., & Makris, A. M. (2023). Identification and Construction of Strong Promoters in Yarrowia lipolytica Suitable for Glycerol-Based Bioprocesses. *Microorganisms*, *11*(5), 1152. https://doi.org/10.3390/MICROORGANISMS11051152

GFI. (2025). *Precision fermentation nomenclature and messaging*.

Ghag, S. B., Adki, V. S., Ganapathi, T. R., & Bapat, V. A. (2021). Plant Platforms for Efficient Heterologous Protein Production. *Biotechnology and Bioprocess Engineering 2021 26:4*, *26*(4), 546–567. https://doi.org/10.1007/S12257-020-0374-1

Gharios, R., Ross, M. L., Li, A., Kottantharayil, S. P., Hoye, J., & DeForest, C. A. (2025). Boolean logic-gated protein presentation through autonomously compiled molecular topology. *Nature Chemical Biology 2025*, 1–11. https://doi.org/10.1038/S41589-025-02037-5

Gibson, D. G., Benders, G. A., Andrews-Pfannkoch, C., Denisova, E. A., Baden-Tillson, H., Zaveri, J., Stockwell, T. B., Brownley, A., Thomas, D. W., Algire, M. A., Merryman, C., Young, L., Noskov, V. N., Glass, J. I., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2008). Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. *Science*, *319*(5867), 1215–1220. https://doi.org/10.1126/SCIENCE.1151721

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R. Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z. N., Segall-Shapiro, T. H., … Venter, J. C. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, *329*(5987), 52–56. https://doi.org/10.1126/SCIENCE.1190719

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods 2009 6:5*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Glass, J. I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M. R., Maruf, M., Hutchison, C. A., Smith, H. O., & Venter, J. C. (2006). Essential genes of a minimal bacterium. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 425–430. https://doi.org/10.1073/PNAS.0510013103

Goeddel, D. V, Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirosef, T., Kraszewskit, A., Itakuraf, K., & Riggstt, A. D. (1979). Expression in Escherichia coli of chemically synthesized genes for human insulin. *Proceedings of the National Academy of Sciences*, *76*(1), 106–110. https://doi.org/10.1073/PNAS.76.1.106

Gong, X., Zhang, J., Gan, Q., Ruan, W., Liu, T., & Yan, Y. (2026). Leveraging artificial intelligence for efficient microbial production. *Bioresource Technology*, *439*, 133286. https://doi.org/10.1016/J.BIORTECH.2025.133286

Gong, Z., Nielsen, J., & Zhou, Y. J. (2017). Engineering Robustness of Microbial Cell Factories. *Biotechnology Journal*, *12*(10), 1700014. https://doi.org/10.1002/BIOT.201700014

Grubbe, W. S., Rasor, B. J., Krüger, A., Jewett, M. C., & Karim, A. S. (2020). Cell-free styrene biosynthesis at high titers. *Metabolic Engineering*, *61*, 89–95. https://doi.org/10.1016/J.YMBEN.2020.05.009

Gülck, T., Booth, J. K., Carvalho, Khakimov, B., Crocoll, C., Motawia, M. S., Møller, B. L., Bohlmann, J., & Gallage, N. J. (2020). Synthetic Biology of Cannabinoids and Cannabinoid Glucosides in Nicotiana benthamiana and Saccharomyces cerevisiae. *Journal of Natural Products*, *83*(10), 2877–2893. https://doi.org/10.1021/ACS.JNATPROD.0C00241

Gutiérrez, S., Overmans, S., Wellman, G. B., Samaras, V. G., Oviedo, C., Gede, M., Szekely, G., & Lauersen, K. J. (2024). A synthetic biology and green bioprocess approach to recreate agarwood sesquiterpenoid mixtures. *Green Chemistry*, *26*(5), 2577–2591. https://doi.org/10.1039/D3GC03708H

Hayashi, N., Lai, Y., Fuerte-Stone, J., Mimee, M., & Lu, T. K. (2024). Cas9-assisted biological containment of a genetically engineered human commensal bacterium and genetic elements. *Nature Communications 2024 15:1*, *15*(1), 1–17. https://doi.org/10.1038/s41467-024-45893-w

Helm, T., Stausberg, T., Previati, M., Ernst, P., Klein, B., Busche, T., Kalinowski, J., Wibberg, D., Wiechert, W., Claerhout, L., Wierckx, N., & Noack, S. (2024). Itaconate Production from Crude Substrates with U. maydis: Scale-up of an Industrially Relevant Bioprocess. *Microbial Cell Factories*, *23*(1), 1–14. https://doi.org/10.1186/S12934-024-02295-3

Henry, L. P., & Bergelson, J. (2025). Applying ecological principles to microbiome engineering. *Nature Microbiology 2025*, 1–11. https://doi.org/10.1038/s41564-025-02076-7

Hettinga, K., & Bijl, E. (2022). Can recombinant milk proteins replace those produced by animals? *Current Opinion in Biotechnology*, *75*, 102690. https://doi.org/10.1016/J.COPBIO.2022.102690

Himel, M. H., Sikder, B., Ahmed, T., & Choudhury, S. M. (2023). Biomimicry in nanotechnology: a comprehensive review. *Nanoscale Advances*, *5*(3), 596–614. https://doi.org/10.1039/D2NA00571A

Hiraizumi, M., Perry, N. T., Durrant, M. G., Soma, T., Nagahata, N., Okazaki, S., Athukoralage, J. S., Isayama, Y., Pai, J. J., Pawluk, A., Konermann, S., Yamashita, K., Hsu, P. D., & Nishimasu, H. (2024). Structural mechanism of bridge RNA-guided recombination. *Nature 2024 630:8018*, *630*(8018), 994–1002. https://doi.org/10.1038/s41586-024-07570-2

Hocquette, J. F., Chriki, S., Fournier, D., & Ellies-Oury, M. P. (2025). Review: Will “cultured meat” transform our food system towards more sustainability? *Animal*, *19*, 101145. https://doi.org/10.1016/J.ANIMAL.2024.101145

Hodgman, C. E., & Jewett, M. C. (2012). Cell-free synthetic biology: Thinking outside the cell. *Metabolic Engineering*, *14*(3), 261–269. https://doi.org/10.1016/J.YMBEN.2011.09.002

Hong, M., Li, T., Xue, W., Zhang, S., Cui, L., Wang, H., Zhang, Y., Zhou, L., Gu, Y., Xia, N., & Li, S. (2022). Genetic engineering of baculovirus-insect cell system to improve protein production. *Frontiers in Bioengineering and Biotechnology*, *10*, 994743. https://doi.org/10.3389/FBIOE.2022.994743

Hoose, A., Vellacott, R., Storch, M., Freemont, P. S., & Ryadnov, M. G. (2023). DNA synthesis technologies to close the gene writing gap. *Nature Reviews Chemistry 2023 7:3*, *7*(3), 144–161. https://doi.org/10.1038/s41570-022-00456-9

Hopes, A., Nekrasov, V., Kamoun, S., & Mock, T. (2016). Editing of the urease gene by CRISPR-Cas in the diatom Thalassiosira pseudonana. *Plant Methods*, *12*(1), 1–12. https://doi.org/10.1186/S13007-016-0148-0

Huang, J., Liu, Z., Bloomer, B. J., Clark, D. S., Mukhopadhyay, A., Keasling, J. D., & Hartwig, J. F. (2021). Unnatural biosynthesis by an engineered microorganism with heterologously expressed natural enzymes and an artificial metalloenzyme. *Nature Chemistry*, *13*(12), 1186–1191. https://doi.org/10.1038/S41557-021-00801-3

Huang, Q., Roessner, C. A., Croteau, R., & Scott, A. I. (2001). Engineering Escherichia coli for the synthesis of taxadiene, a key intermediate in the biosynthesis of taxol. *Bioorganic & Medicinal Chemistry*, *9*(9), 2237–2242. https://doi.org/10.1016/S0968-0896(01)00072-4

Huang, X., Wang, Y., & Wang, N. (2022). Highly Efficient Generation of Canker-Resistant Sweet Orange Enabled by an Improved CRISPR/Cas9 System. *Frontiers in Plant Science*, *12*, 769907. https://doi.org/10.3389/FPLS.2021.769907

Humbird, D. (2021). Scale-up economics for cultured meat. *Biotechnology and Bioengineering*, *118*(8), 3239–3250. https://doi.org/10.1002/BIT.27848

Jansson, J. K., McClure, R., & Egbert, R. G. (2023). Soil microbiome engineering for sustainability in a changing environment. *Nature Biotechnology 2023 41:12*, *41*(12), 1716–1728. https://doi.org/10.1038/s41587-023-01932-3

Järviö, N., Parviainen, T., Maljanen, N. L., Kobayashi, Y., Kujanpää, L., Ercili-Cura, D., Landowski, C. P., Ryynänen, T., Nordlund, E., & Tuomisto, H. L. (2021). Ovalbumin production using Trichoderma reesei culture and low-carbon energy could mitigate

the environmental impacts of chicken-egg-derived ovalbumin. *Nature Food 2022 2:12*, *2*(12), 1005–1013. https://doi.org/10.1038/s43016-021-00418-2

Jia, M., Zhu, S., Xue, M. Y., Chen, H., Xu, J., Song, M., Tang, Y., Liu, X., Tao, Y., Zhang, T., Liu, J. X., Wang, Y., & Sun, H. Z. (2024). Single-cell transcriptomics across 2,534 microbial species reveals functional heterogeneity in the rumen microbiome. *Nature Microbiology 2024 9:7*, *9*(7), 1884–1898. https://doi.org/10.1038/s41564-024-01723-9

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816–821. https://doi.org/10.1126/SCIENCE.1225829

Johnson, N. W., Valenzuela-Ortega, M., Thorpe, T. W., Era, Y., Kjeldsen, A., Mulholland, K., & Wallace, S. (2025). A biocompatible Lossen rearrangement in Escherichia coli. *Nature Chemistry 2025*, 1–7. https://doi.org/10.1038/s41557-025-01845-5

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. https://doi.org/10.1038/S41586-021-03819-2

Jung, V., Vincent-Cuaz, C., Tumescheit, C., Fournier, L., Darsinou, M., Xu, Z. M., Saadat, A., Wang, Y., Tsantoulis, P., Michielin, O., Fellay, J., Patani, R., Ramos, A., Frossard, P., Hastings, J., Riccio, A., van der Plas, L., & Luisier, R. (2025). Decoding the interactions and functions of non-coding RNA with artificial intelligence. *Nature Reviews Molecular Cell Biology 2025*, 1–22. https://doi.org/10.1038/s41580-025-00857-w

Kabadi, A. M., Ousterout, D. G., Hilton, I. B., & Gersbach, C. A. (2014). Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Research*, *42*(19), e147–e147. https://doi.org/10.1093/NAR/GKU749

Kampers, L. F. C., Asin-Garcia, E., Schaap, P. J., Wagemakers, A., & Martins dos Santos, V. A. P. (2021). From Innovation to Application: Bridging the Valley of Death in Industrial

Biotechnology. *Trends in Biotechnology*, *39*(12), 1240–1242. https://doi.org/10.1016/j.tibtech.2021.04.010

Kay, J. E., & Jewett, M. C. (2020). A cell-free system for production of 2,3-butanediol is robust to growth-toxic compounds. *Metabolic Engineering Communications*, *10*, e00114. https://doi.org/10.1016/J.MEC.2019.E00114

Ke, J., Wang, B., & Yoshikuni, Y. (2021). Microbiome Engineering: Synthetic Biology of Plant-Associated Microbiomes in Sustainable Agriculture. *Trends in Biotechnology*, *39*(3), 244–261. https://doi.org/10.1016/J.TIBTECH.2020.07.008

Keasling, J., Garcia Martin, H., Lee, T. S., Mukhopadhyay, A., Singer, S. W., & Sundstrom, E. (2021). Microbial production of advanced biofuels. *Nature Reviews Microbiology 2021 19:11*, *19*(11), 701–715. https://doi.org/10.1038/s41579-021-00577-w

Khalil, A. S., & Collins, J. J. (2010). Synthetic biology: applications come of age. *Nature Reviews Genetics 2010 11:5*, *11*(5), 367–379. https://doi.org/10.1038/nrg2775

Kim, K. H., Chandran, D., & Sauro, H. M. (2011). Toward Modularity in Synthetic Biology: Design Patterns and Fan-out. *Design and Analysis of Biomolecular Circuits*, 117–138. https://doi.org/10.1007/978-1-4419-6766-4_6

Kim, N., Ma, J., Kim, W., Kim, J., Belenky, P., & Lee, I. (2024). Genome-resolved metagenomics: a game changer for microbiome medicine. *Experimental & Molecular Medicine 2024 56:7*, *56*(7), 1501–1512. https://doi.org/10.1038/s12276-024-01262-7

Kittler, S., Müller, F., Elshazly, M., Wandrey, G. B., Klein, T., Daub, A., Spadiut, O., & Kopp, J. (2025). Transferability of bioprocessing modes for recombinant protease production: from fed-batch to continuous cultivation with Bacillus licheniformis. *BMC Biotechnology*, *25*(1), 1–13. https://doi.org/10.1186/S12896-025-00947-9

Knight, T. (2003). *Idempotent Vector Design for Standard Assembly of Biobricks*. https://dspace.mit.edu/handle/1721.1/21168

Kobayashi, Y., Inokuma, K., Matsuda, M., Kondo, A., & Hasunuma, T. (2021). Resveratrol production from several types of saccharide sources by a recombinant

Scheffersomyces stipitis strain. *Metabolic Engineering Communications*, *13*, e00188. https://doi.org/10.1016/J.MEC.2021.E00188

Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature 2015 533:7603*, *533*(7603), 420–424. https://doi.org/10.1038/nature17946

Koralesky, K. E., Sirovica, L. V., Hendricks, J., Mills, K. E., von Keyserlingk, M. A. G., & Weary, D. M. (2023). Social acceptance of genetic engineering technology. *PLOS ONE*, *18*(8), e0290070. https://doi.org/10.1371/JOURNAL.PONE.0290070

Korot, E., Pontikos, N., Liu, X., Wagner, S. K., Faes, L., Huemer, J., Balaskas, K., Denniston, A. K., Khawaja, A., & Keane, P. A. (2021). Predicting sex from retinal fundus photographs using automated deep learning. *Scientific Reports*, *11*(1), 1–8. https://doi.org/10.1038/S41598-021-89743-X

Kortemme, T. (2024). De novo protein design—From new structures to programmable functions. *Cell*, *187*(3), 526–544. https://doi.org/10.1016/J.CELL.2023.12.028

Kosuri, S., & Church, G. M. (2014). Large-scale de novo DNA synthesis: technologies and applications. *Nature Methods 2014 11:5*, *11*(5), 499–507. https://doi.org/10.1038/nmeth.2918

Kouprina, N., & Larionov, V. (2006). TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nature Reviews Genetics 2006 7:10*, *7*(10), 805–812. https://doi.org/10.1038/nrg1943

Kuivanen, J., Wang, Y. M. J., & Richard, P. (2016). Engineering Aspergillus niger for galactaric acid production: Elimination of galactaric acid catabolism by using RNA sequencing and CRISPR/Cas9. *Microbial Cell Factories*, *15*(1), 1–9. https://doi.org/10.1186/S12934-016-0613-5

Kumar, S., Verma, N., & Singh, A. K. (2017). Development of cadmium specific recombinant biosensor and its application in milk samples. *Sensors and Actuators B: Chemical*, *240*, 248–254. https://doi.org/10.1016/J.SNB.2016.08.160

Lam, E., & Michael, T. P. (2022). Wolffia, a minimalist plant and synthetic biology chassis. *Trends in Plant Science*, *27*(5), 430–439. https://doi.org/10.1016/j.tplants.2021.11.014

Lambert, E. G., O'Keeffe, C. J., Ward, A. O., Anderson, T. A., Yip, Q., & Newman, P. L. H. (2024). Enhancing the palatability of cultivated meat. *Trends in Biotechnology*, *42*(9), 1112–1127. https://doi.org/10.1016/j.tibtech.2024.02.014

Lanz, M., Hartmann, C., Egan, P., & Siegrist, M. (2024). Consumer acceptance of cultured, plant-based, 3D-printed meat and fish alternatives. *Future Foods*, *9*, 100297. https://doi.org/10.1016/J.FUFO.2024.100297

Larroude, M., Celinska, E., Back, A., Thomas, S., Nicaud, J. M., & Ledesma-Amaro, R. (2018). A synthetic biology approach to transform Yarrowia lipolytica into a competitive biotechnological producer of β-carotene. *Biotechnology and Bioengineering*, *115*(2), 464–472. https://doi.org/10.1002/BIT.26473

Lartigue, C., Glass, J. I., Alperovich, N., Pieper, R., Parmar, P. P., Hutchison, C. A., Smith, H. O., & Venter, J. C. (2007). Genome transplantation in bacteria: Changing one species to another. *Science*, *317*(5838), 632–638. https://doi.org/10.1126/SCIENCE.1144622

Lee, D. H., Kim, H., Sung, B. H., Cho, B. K., & Lee, S. G. (2023). Biofoundries: Bridging Automation and Biomanufacturing in Synthetic Biology. *Biotechnology and Bioprocess Engineering*, *28*(6), 892–904. https://doi.org/10.1007/S12257-023-0226-X

Lee, M. E., DeLoache, W. C., Cervantes, B., & Dueber, J. E. (2015). A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology*, *4*(9), 975–986. https://doi.org/10.1021/SB500366V

Lee, S. Y., Kim, H. U., Chae, T. U., Cho, J. S., Kim, J. W., Shin, J. H., Kim, D. I., Ko, Y. S., Jang, W. D., & Jang, Y. S. (2019). A comprehensive metabolic map for production of bio-based chemicals. *Nature Catalysis 2019 2:1*, *2*(1), 18–33. https://doi.org/10.1038/s41929-018-0212-4

Lei, Q., Ma, J., Du, G., Zhou, J., & Guan, X. (2023). Efficient expression of a cytokine combination in Saccharomyces cerevisiae for cultured meat production. *Food*

Research International, *170*, 113017. https://doi.org/10.1016/J.FOODRES.2023.113017

Leong, H. Y., Chang, C. K., Khoo, K. S., Chew, K. W., Chia, S. R., Lim, J. W., Chang, J. S., & Show, P. L. (2021). Waste biorefinery towards a sustainable circular bioeconomy: a solution to global issues. *Biotechnology for Biofuels*, *14*(1), 1–15. https://doi.org/10.1186/S13068-021-01939-5

Leprince, A., van Passel, M. W. J., & dos Santos, V. A. P. M. (2012). Streamlining genomes: toward the generation of simplified and stabilized microbial systems. *Current Opinion in Biotechnology*, *23*(5), 651–658. https://doi.org/10.1016/J.COPBIO.2012.05.001

Li, J., Meng, X., Zong, Y., Chen, K., Zhang, H., Liu, J., Li, J., & Gao, C. (2016). Gene replacements and insertions in rice by intron targeting using CRISPR–Cas9. *Nature Plants 2016 2:10*, *2*(10), 1–6. https://doi.org/10.1038/nplants.2016.139

Li, M., Chen, R., Jiang, Q., Sun, X., Zhang, H., & Hu, Z. (2021). GmNAC06, a NAC domain transcription factor enhances salt stress tolerance in soybean. *Plant Molecular Biology*, *105*(3), 333–345. https://doi.org/10.1007/S11103-020-01091-Y

Li, R., Li, R., Li, X., Fu, D., Zhu, B., Tian, H., Luo, Y., & Zhu, H. (2018). Multiplexed CRISPR/Cas9-mediated metabolic engineering of γ-aminobutyric acid levels in Solanum lycopersicum. *Plant Biotechnology Journal*, *16*(2), 415–427. https://doi.org/10.1111/PBI.12781

Liaquat, F., Munis, M. F. H., Arif, S., Haroon, U., Shi, J., Saqib, S., Zaman, W., Che, S., & Liu, Q. (2021). PacBio Single-Molecule Long-Read Sequencing Reveals Genes Tolerating Manganese Stress in Schima superba Saplings. *Frontiers in Genetics*, *12*, 635043. https://doi.org/10.3389/FGENE.2021.635043

Liew, F. E., Nogle, R., Abdalla, T., Rasor, B. J., Canter, C., Jensen, R. O., Wang, L., Strutz, J., Chirania, P., De Tissera, S., Mueller, A. P., Ruan, Z., Gao, A., Tran, L., Engle, N. L., Bromley, J. C., Daniell, J., Conrado, R., Tschaplinski, T. J., … Köpke, M. (2022). Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. *Nature Biotechnology 2022 40:3*, *40*(3), 335–344. https://doi.org/10.1038/s41587-021-01195-w

Listov, D., Goverde, C. A., Correia, B. E., & Fleishman, S. J. (2024). Opportunities and challenges in design and optimization of protein function. *Nature Reviews Molecular Cell Biology*, *25*(8), 639–653. https://doi.org/10.1038/S41580-024-00718-Y

Liu, A. P., Appel, E. A., Ashby, P. D., Baker, B. M., Franco, E., Gu, L., Haynes, K., Joshi, N. S., Kloxin, A. M., Kouwer, P. H. J., Mittal, J., Morsut, L., Noireaux, V., Parekh, S., Schulman, R., Tang, S. K. Y., Valentine, M. T., Vega, S. L., Weber, W., … Chaudhuri, O. (2022). The living interface between synthetic biology and biomaterial design. *Nature Materials*, *21*(4), 390–397. https://doi.org/10.1038/S41563-022-01231-3

Liu, C. P., Tsai, T. I., Cheng, T., Shivatare, V. S., Wu, C. Y., Wu, C. Y., & Wong, C. H. (2018). Glycoengineering of antibody (Herceptin) through yeast expression and in vitro enzymatic glycosylation. *Proceedings of the National Academy of Sciences*, *115*(4), 720–725. https://doi.org/10.1073/PNAS.1718172115

Liu, D., Garrigues, S., & de Vries, R. P. (2023). Heterologous protein production in filamentous fungi. *Applied Microbiology and Biotechnology 2023 107:16*, *107*(16), 5019–5033. https://doi.org/10.1007/S00253-023-12660-8

Loman, N. J., Quick, J., & Simpson, J. T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods 2015 12:8*, *12*(8), 733–735. https://doi.org/10.1038/nmeth.3444

Lopatkin, A. J., & Collins, J. J. (2020). Predictive biology: modelling, understanding and harnessing microbial complexity. *Nature Reviews Microbiology 2020 18:9*, *18*(9), 507–520. https://doi.org/10.1038/s41579-020-0372-5

Lorenzo, V. de, Prather, K. L., Chen, G., O'Day, E., Kameke, C. von, Oyarzún, D. A., Hosta-Rigau, L., Alsafar, H., Cao, C., Ji, W., Okano, H., Roberts, R. J., Ronaghi, M., Yeung, K., Zhang, F., & Lee, S. Y. (2018). The power of synthetic biology for bioproduction, remediation and pollution control. *EMBO Reports*, *19*(4), 21. https://doi.org/10.15252/EMBR.201745658

Lowder, L. G., Zhang, D., Baltes, N. J., Paul, J. W., Tang, X., Zheng, X., Voytas, D. F., Hsieh, T. F., Zhang, Y., & Qi, Y. (2015). A CRISPR/Cas9 Toolbox for Multiplexed Plant Genome

Editing and Transcriptional Regulation. *Plant Physiology*, *169*(2), 971–985. https://doi.org/10.1104/PP.15.00636

Lu, H., Xiao, L., Liao, W., Yan, X., & Nielsen, J. (2024). Cell factory design with advanced metabolic modelling empowered by artificial intelligence. *Metabolic Engineering*, *85*, 61–72. https://doi.org/10.1016/J.YMBEN.2024.07.003

Luo, X., Reiter, M. A., d'Espaux, L., Wong, J., Denby, C. M., Lechner, A., Zhang, Y., Grzybowski, A. T., Harth, S., Lin, W., Lee, H., Yu, C., Shin, J., Deng, K., Benites, V. T., Wang, G., Baidoo, E. E. K., Chen, Y., Dev, I., … Keasling, J. D. (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature 2019 567:7746*, *567*(7746), 123–126. https://doi.org/10.1038/s41586-019-0978-9

Ma, H., Marti-Gutierrez, N., Park, S. W., Wu, J., Lee, Y., Suzuki, K., Koski, A., Ji, D., Hayama, T., Ahmed, R., Darby, H., Van Dyken, C., Li, Y., Kang, E., Park, A. R., Kim, D., Kim, S. T., Gong, J., Gu, Y., … Mitalipov, S. (2017). Correction of a pathogenic gene mutation in human embryos. *Nature 2017 548:7668*, *548*(7668), 413–419. https://doi.org/10.1038/nature23305

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, *339*(6121), 823–826. https://doi.org/10.1126/SCIENCE.1232033

Mandakovic, D., Cintolesi, Á., Maldonado, J., Mendoza, S. N., Aïte, M., Gaete, A., Saitua, F., Allende, M., Cambiazo, V., Siegel, A., Maass, A., González, M., & Latorre, M. (2020). Genome-scale metabolic models of Microbacterium species isolated from a high altitude desert environment. *Scientific Reports 2020 10:1*, *10*(1), 1–13. https://doi.org/10.1038/s41598-020-62130-8

Mandell, D. J., Lajoie, M. J., Mee, M. T., Takeuchi, R., Kuznetsov, G., Norville, J. E., Gregg, C. J., Stoddard, B. L., & Church, G. M. (2015). Biocontainment of genetically modified organisms by synthetic protein design. *Nature 2015 518:7537*, *518*(7537), 55–60. https://doi.org/10.1038/nature14121

Manduva, V. C. (2023). Unlocking Growth Potential at the Intersection of AI, Robotics, and Synthetic Biology. *International Journal of Modern Computing*, *6*(1), 53–63. https://yuktabpublisher.com/index.php/IJMC/article/view/123

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., … Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature 2005 437:7057*, *437*(7057), 376–380. https://doi.org/10.1038/nature03959

Mateos Fernández, R., Petek, M., Gerasymenko, I., Juteršek, M., Baebler, Š., Kallam, K., Moreno Giménez, E., Gondolf, J., Nordmann, A., Gruden, K., Orzaez, D., & Patron, N. J. (2022). Insect pest management in the age of synthetic biology. *Plant Biotechnology Journal*, *20*(1), 25–36. https://doi.org/10.1111/PBI.13685

McCarty, N. S., & Ledesma-Amaro, R. (2019). Synthetic Biology Tools to Engineer Microbial Communities for Biotechnology. *Trends in Biotechnology*, *37*(2), 181–197. https://doi.org/10.1016/J.TIBTECH.2018.11.002

Meng, F., & Ellis, T. (2020). The second decade of synthetic biology: 2010–2020. *Nature Communications 2020 11:1*, *11*(1), 1–4. https://doi.org/10.1038/s41467-020-19092-2

Mirsalami, S. M., & Mirsalami, M. (2025). Advances in genetically engineered microorganisms: Transforming food production through precision fermentation and synthetic biology. *Future Foods*, *11*, 100601. https://doi.org/10.1016/J.FUFO.2025.100601

Mitchell, L. A., Wang, A., Stracquadanio, G., Kuang, Z., Wang, X., Yang, K., Richardson, S., Martin, J. A., Zhao, Y., Walker, R., Luo, Y., Dai, H., Dong, K., Tang, Z., Yang, Y., Cai, Y., Heguy, A., Ueberheide, B., Fenyö, D., … Boeke, J. D. (2017). Synthesis, debugging, and effects of synthetic chromosome consolidation: synVI and beyond. *Science*, *355*(6329). https://doi.org/10.1126/SCIENCE.AAF4831

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic

elements. *Journal of Molecular Evolution*, *60*(2), 174–182. https://doi.org/10.1007/S00239-004-0046-3

More, S., Bampidis, V., Benford, D., Bragard, C., Halldorsson, T., Hernández-Jerez, A., Bennekou, S. H., Koutsoumanis, K., Lambré, C., Machera, K., Mullins, E., Nielsen, S. S., Schlatter, J., Schrenk, D., Turck, D., Younes, M., Herman, L., Pelaez, C., van Loveren, H., … Cocconcelli, P. S. (2022). Evaluation of existing guidelines for their adequacy for the food and feed risk assessment of microorganisms obtained through synthetic biology. *EFSA Journal*, *20*(8), e07479. https://doi.org/10.2903/J.EFSA.2022.7479

Mousa, W. K., Ghemrawi, R., Abu-Izneid, T., Al Ramadan, N., & Al Sheebani, F. (2024). The design and development of EcoBiomes: Multi-species synthetic microbial consortia inspired by natural desert microbiome to enhance the resilience of climate-sensitive ecosystems. *Heliyon*, *10*(16). https://doi.org/10.1016/J.HELIYON.2024.E36548

Mózsik, L., Pohl, C., Meyer, V., Bovenberg, R. A. L., Nygård, Y., & Driessen, A. J. M. (2021). Modular Synthetic Biology Toolkit for Filamentous Fungi. *ACS Synthetic Biology*, *10*(11), 2850–2861. https://doi.org/10.1021/ACSSYNBIO.1C00260

Mu'azzam, K., Santos da Silva, F. V., Murtagh, J., & Sousa Gallagher, M. J. (2024). A roadmap for model-based bioprocess development. *Biotechnology Advances*, *73*, 108378. https://doi.org/10.1016/J.BIOTECHADV.2024.108378

Mullins, E., Bresson, J. L., Dewhurst, I. C., Epstein, M. M., Firbank, L. G., Guerche, P., Hejatko, J., Moreno, F. J., Naegeli, H., Nogué, F., Rostoks, N., Sánchez Serrano, J. J., Savoini, G., Veromann, E., Veronesi, F., Cocconcelli, P. S., Glandorf, D., Herman, L., Jimenez Saiz, R., … Dalmay, T. (2024). New developments in biotechnology applied to microorganisms. *EFSA Journal*, *22*(7), e8895. https://doi.org/10.2903/J.EFSA.2024.8895

Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross, D., Densmore, D., & Voigt, C. A. (2016). Genetic circuit design automation. *Science*, *352*(6281). https://doi.org/10.1126/SCIENCE.AAC7341

Nielsen, J., & Keasling, J. D. (2016). Engineering Cellular Metabolism. *Cell*, *164*(6), 1185–1197. https://doi.org/10.1016/J.CELL.2016.02.004

Olsson, L., Rugbjerg, P., Torello Pianale, L., & Trivellin, C. (2022). Robustness: linking strain design to viable bioprocesses. *Trends in Biotechnology*, *40*(8), 918–931. https://doi.org/10.1016/J.TIBTECH.2022.01.004

Olynk Widmar, N. J., Dominick, S. R., Tyner, W. E., & Ruple, A. (2017). When is genetic modification socially acceptable? When used to advance human health through avenues other than food. *PLOS ONE*, *12*(6), e0178227. https://doi.org/10.1371/JOURNAL.PONE.0178227

O'Malley, M. A., Powell, A., Davies, J. F., & Calvert, J. (2008). Knowledge-making distinctions in synthetic biology. *BioEssays*, *30*(1), 57–65. https://doi.org/10.1002/BIES.20664

Ostrov, N., Landon, M., Guell, M., Kuznetsov, G., Teramoto, J., Cervantes, N., Zhou, M., Singh, K., Napolitano, M. G., Moosburner, M., Shrock, E., Pruitt, B. W., Conway, N., Goodman, D. B., Gardner, C. L., Tyree, G., Gonzales, A., Wanner, B. L., Norville, J. E., … Church, G. M. (2016). Design, synthesis, and testing toward a 57-codon genome. *Science (New York, N.Y.)*, *353*(6301), 819–822. https://doi.org/10.1126/SCIENCE.AAF3639

Ousterout, D. G., Kabadi, A. M., Thakore, P. I., Majoros, W. H., Reddy, T. E., & Gersbach, C. A. (2015). Multiplex CRISPR/Cas9-based genome editing for correction of dystrophin mutations that cause Duchenne muscular dystrophy. *Nature Communications 2015 6:1*, *6*(1), 1–13. https://doi.org/10.1038/ncomms7244

Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M. D., Tai, A., Main, A., Eng, D., Polichuk, D. R., Teoh, K. H., Reed, D. W., Treynor, T., Lenihan, J., Jiang, H., Fleck, M., Bajad, S., Dang, G., … Newman, J. D. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature 2013 496:7446*, *496*(7446), 528–532. https://doi.org/10.1038/nature12051

Palluk, S., Arlow, D. H., De Rond, T., Barthel, S., Kang, J. S., Bector, R., Baghdassarian, H. M., Truong, A. N., Kim, P. W., Singh, A. K., Hillson, N. J., & Keasling, J. D. (2018). De novo DNA synthesis using polymerasenucleotide conjugates. *Nature Biotechnology*, *36*(7), 645–650. https://doi.org/10.1038/NBT.4173

Palumbi, S. R. (2001). Humans as the world's greatest evolutionary force. *Science*, *293*(5536), 1786–1790. https://doi.org/10.1126/SCIENCE.293.5536.1786

Pan, X., Tang, M., You, J., Hao, Y., Zhang, X., Yang, T., & Rao, Z. (2023). A Novel Method to Screen Strong Constitutive Promoters in Escherichia coli and Serratia marcescens for Industrial Applications. *Biology*, *12*(1), 71. https://doi.org/10.3390/BIOLOGY12010071

Paudel, B., Kolady, D. E., Just, D., & Van der Sluis, E. (2023). Determinants of consumer acceptance of gene-edited foods and its implications for innovators and policymakers. *Agribusiness*, *39*(3), 623–645. https://doi.org/10.1002/AGR.21799

Peng, H., Chen, R., Shaw, W. M., Hapeta, P., Jiang, W., Bell, D. J., Ellis, T., & Ledesma-Amaro, R. (2023). Modular Metabolic Engineering and Synthetic Coculture Strategies for the Production of Aromatic Compounds in Yeast. *ACS Synthetic Biology*, *12*(6), 1739–1749. https://doi.org/10.1021/ACSSYNBIO.3C00047

Poppeliers, J., Boon, M., De Mey, M., Masschelein, J., & Lavigne, R. (2025). Non-model bacteria as platforms for endogenous gene expression in synthetic biology. *Nature Reviews Bioengineering 2025*, 1–15. https://doi.org/10.1038/S44222-025-00354-X

Port, F., & Bullock, S. L. (2016). Augmenting CRISPR applications in Drosophila with tRNA-flanked sgRNAs. *Nature Methods 2016 13:10*, *13*(10), 852–854. https://doi.org/10.1038/nmeth.3972

Post, M. J. (2012). Cultured meat from stem cells: Challenges and prospects. *Meat Science*, *92*(3), 297–301. https://doi.org/10.1016/J.MEATSCI.2012.04.008

Purnick, P. E. M., & Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology 2009 10:6*, *10*(6), 410–422. https://doi.org/10.1038/nrm2698

Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell*, *152*(5), 1173–1183. https://doi.org/10.1016/j.cell.2013.02.022

Qu, Y., Huang, K., Yin, M., Zhan, K., Liu, D., Yin, D., Cousins, H. C., Johnson, W. A., Wang, X., Shah, M., Altman, R. B., Zhou, D., Wang, M., & Cong, L. (2025). CRISPR-GPT for agentic automation of gene-editing experiments. *Nature Biomedical Engineering 2025*, 1–14. https://doi.org/10.1038/s41551-025-01463-z

Rasul, M. G., Hazrat, M. A., Sattar, M. A., Jahirul, M. I., & Shearer, M. J. (2022). The future of hydrogen: Challenges on production, storage and applications. *Energy Conversion and Management*, *272*, 116326. https://doi.org/10.1016/J.ENCONMAN.2022.116326

Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., Karchner, S. I., Hahn, M. E., Nacci, D., Oleksiak, M. F., Crawford, D. L., & Whitehead, A. (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science*, *354*(6317), 1305–1308. https://doi.org/10.1126/SCIENCE.AAH4993

Reiss, J., Robertson, S., & Suzuki, M. (2021). Cell Sources for Cultivated Meat: Applications and Considerations throughout the Production Workflow. *International Journal of Molecular Sciences 2021, Vol. 22, Page 7513*, *22*(14), 7513. https://doi.org/10.3390/IJMS22147513

Ren, H., Qin, M., Zhang, L., Li, Z., Li, Y., He, Q., Zhong, J., Zhao, D., Lian, X., Jiang, H., Liao, X., & Sun, J. (2025). Modular Engineering of a Synthetic Biology-Based Platform for Sustainable Bioremediation of Residual Antibiotics in Aquatic Environments. *Engineering*. https://doi.org/10.1016/J.ENG.2025.03.033

Restrepo-Pérez, L., Wong, C. H., Maglia, G., Dekker, C., & Joo, C. (2019). Label-Free Detection of Post-translational Modifications with a Nanopore. *Nano Letters*, *19*(11), 7957–7964. https://doi.org/10.1021/ACS.NANOLETT.9B03134

Rice, A. J., Sword, T. T., Chengan, K., Mitchell, D. A., Mouncey, N. J., Moore, S. J., & Bailey, C. B. (2025). Cell-free synthetic biology for natural product biosynthesis and discovery. *Chemical Society Reviews*, *54*(9), 4314–4352. https://doi.org/10.1039/D4CS01198H

Ro, D. K., Paradise, E. M., Quellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C. Y., Withers, S. T., Shiba, Y., Sarpong, R., & Keasling, J. D. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature 2006 440:7086*, *440*(7086), 940–943. https://doi.org/10.1038/nature04640

Robertson, W. E., Funke, L. F. H., de la Torre, D., Fredens, J., Elliott, T. S., Spinck, M., Christova, Y., Cervettini, D., Böge, F. L., Liu, K. C., Buse, S., Maslen, S., Salmond, G. P. C., & Chin, J. W. (2021). Sense codon reassignment enables viral resistance and encoded polymer synthesis. *Science*, *372*(6546), 1057–1062. https://doi.org/10.1126/SCIENCE.ABG3029

Røkke, G., Korvald, E., Pahr, J., Øyås, O., & Lale, R. (2014). BioBrick Assembly Standards and Techniques and Associated Software Tools. *Methods in Molecular Biology*, *1116*, 1–24. https://doi.org/10.1007/978-1-62703-764-8_1

Rönspies, M., Dorn, A., Schindele, P., & Puchta, H. (2021). CRISPR–Cas-mediated chromosome engineering for crop improvement and synthetic biology. *Nature Plants 2021 7:5*, *7*(5), 566–573. https://doi.org/10.1038/s41477-021-00910-4

Rottinghaus, A. G., Ferreiro, A., Fishbein, S. R. S., Dantas, G., & Moon, T. S. (2022). Genetically stable CRISPR-based kill switches for engineered microbes. *Nature Communications 2022 13:1*, *13*(1), 1–17. https://doi.org/10.1038/s41467-022-28163-5

Ruan, Z., Xu, M., Xing, Y., Yang, K., Xu, X., Jiang, J., & Qiu, R. (2024). Enhanced growth of wheat in contaminated fields via synthetic microbiome as revealed by genome-scale metabolic modeling. *Science of The Total Environment*, *953*, 176047. https://doi.org/10.1016/J.SCITOTENV.2024.176047

Sebé-Pedrós, A., Tanay, A., Lawniczak, M. K. N., Arendt, D., Aerts, S., Archibald, J., Arnone, M. I., Blaxter, M., Cleves, P., Coelho, S. M., Dias, M., Dunn, C., Elek, A., Frazer, J., Gabaldón, T., Gillis, J., Grau-Bové, X., Guigó, R., Hobert, O., … Wang, B. (2025). The Biodiversity Cell Atlas: mapping the tree of life at cellular resolution. *Nature 2025 645:8082*, *645*(8082), 877–885. https://doi.org/10.1038/S41586-025-09312-4

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W. R., Bridgland, A., Penedones, H., Petersen, S., Simonyan, K., Crossan, S., Kohli, P., Jones, D. T., Silver, D., Kavukcuoglu, K., & Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature 2020 577:7792*, *577*(7792), 706–710. https://doi.org/10.1038/s41586-019-1923-7

Sessou, P., Keisam, S., Tuikhar, N., Gagara, M., Farougou, S., & Jeyaram, K. (2019). High-throughput Illumina Miseq amplicon sequencing of yeast communities associated with indigenous dairy products from republics of Benin and Niger. *Frontiers in Microbiology*, *10*(APR), 439311. https://doi.org/10.3389/FMICB.2019.00594

Shabestary, K., Klemm, C., Carling, B., Marshall, J., Savigny, J., Storch, M., & Ledesma-Amaro, R. (2024). Phenotypic heterogeneity follows a growth-viability tradeoff in response to amino acid identity. *Nature Communications 2024 15:1*, *15*(1), 1–16. https://doi.org/10.1038/s41467-024-50602-8

Shaw, W. M., Studená, L., Roy, K., Hapeta, P., McCarty, N. S., Graham, A. E., Ellis, T., & Ledesma-Amaro, R. (2022). Inducible expression of large gRNA arrays for multiplexed CRISPRai applications. *Nature Communications 2022 13:1*, *13*(1), 1–10. https://doi.org/10.1038/s41467-022-32603-7

Shen, Y., Wang, Y., Chen, T., Gao, F., Gong, J., Abramczyk, D., Walker, R., Zhao, H., Chen, S., Liu, W., Luo, Y., Müller, C. A., Paul-Dubois-Taine, A., Alver, B., Stracquadanio, G., Mitchell, L. A., Luo, Z., Fan, Y., Zhou, B., … Yang, H. (2017). Deep functional analysis of synII, a 770-kilobase synthetic yeast chromosome. *Science*, *355*(6329). https://doi.org/10.1126/SCIENCE.AAF4791

Shetty, R. P., Endy, D., & Knight, T. F. (2008). Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering*, *2*(1), 1–12. https://doi.org/10.1186/1754-1611-2-5

Shin, J., Lee, S., Go, M. J., Lee, S. Y., Kim, S. C., Lee, C. H., & Cho, B. K. (2016). Analysis of the mouse gut microbiome using full-length 16S rRNA amplicon sequencing. *Scientific Reports 2016 6:1*, *6*(1), 1–10. https://doi.org/10.1038/srep29681

Shulse, C. N., Chovatia, M., Agosto, C., Yoshikuni, G. Y., Hamilton, M., Deutsch, S., Yoshikuni, Y., & Blow, M. J. (2019). Engineered root bacteria release plant-available phosphate from phytate. *Applied and Environmental Microbiology*, *85*(18). https://doi.org/10.1128/AEM.01210-19

Shumate, A., Wong, B., Pertea, G., & Pertea, M. (2022). Improved transcriptome assembly using a hybrid of long and short reads with StringTie. *PLOS Computational Biology*, *18*(6), e1009730. https://doi.org/10.1371/JOURNAL.PCBI.1009730

Son, H. I., Weiss, A., & You, L. (2021). Design patterns for engineering genetic stability. *Current Opinion in Biomedical Engineering*, *19*, 100297. https://doi.org/10.1016/J.COBME.2021.100297

Soon, W. L., Chong, H. Q., Foo, J. L., & Chang, M. W. (2025). New-to-nature biocompatible chemistry for plastic waste upcycling. *Nature Chemistry 2025*, 1–2. https://doi.org/10.1038/s41557-025-01863-3

South, P. F., Cavanagh, A. P., Liu, H. W., & Ort, D. R. (2019). Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science*, *363*(6422). https://doi.org/10.1126/SCIENCE.AAT9077

Spadiut, O., Rittmann, S., Dietzsch, C., & Herwig, C. (2013). Dynamic process conditions in bioprocess development. *Engineering in Life Sciences*, *13*(1), 88–101. https://doi.org/10.1002/ELSC.201200026

Stephenson, A., Lastra, L., Nguyen, B., Chen, Y. J., Nivala, J., Ceze, L., & Strauss, K. (2023). Physical Laboratory Automation in Synthetic Biology. *ACS Synthetic Biology*, *12*(11), 3156–3169. https://doi.org/10.1021/ACSSYNBIO.3C00345

Sundararajan, P., Ghosh, S., Kelbessa, B. G., Whisson, S. C., Dubey, M., Chawade, A., & Vetukuri, R. R. (2025). The impact of spray-induced gene silencing on cereal phyllosphere microbiota. *Environmental Microbiome 2024 20:1*, *20*(1), 1–17. https://doi.org/10.1186/S40793-024-00660-8

Tan, X., Letendre, J. H., Collins, J. J., & Wong, W. W. (2021). Synthetic biology in the clinic: engineering vaccines, diagnostics, and therapeutics. *Cell*, *184*(4), 881–898. https://doi.org/10.1016/J.CELL.2021.01.017

Tang, T. C., An, B., Huang, Y., Vasikaran, S., Wang, Y., Jiang, X., Lu, T. K., & Zhong, C. (2021). Materials design by synthetic biology. *Nature Reviews Materials*, *6*(4), 332–350. https://doi.org/10.1038/S41578-020-00265-W

Teng, T. S., Chin, Y. L., Chai, K. F., & Chen, W. N. (2021). Fermentation for future food systems. *EMBO Reports*, *22*(5). https://doi.org/10.15252/EMBR.202152680

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology 2022 40:7*, *40*(7), 1023–1025. https://doi.org/10.1038/s41587-021-01156-3

Turnbull, C., Lillemo, M., & Hvoslef-Eide, T. A. K. (2021). Global Regulation of Genetically Modified Crops Amid the Gene Edited Crop Boom – A Review. *Frontiers in Plant Science*, *12*, 630396. https://doi.org/10.3389/FPLS.2021.630396

United Nations. (n.d.-a). *Take Action for the Sustainable Development Goals - United Nations Sustainable Development*. Retrieved December 23, 2025, from https://www.un.org/sustainabledevelopment/sustainable-development-goals/

United Nations. (n.d.-b). *What Is Climate Change?* Retrieved May 28, 2024, from https://www.un.org/en/climatechange/what-is-climate-change

Vaeck, M., Reynaerts, A., Höfte, H., Jansens, S., De Beuckeleer, M., Dean, C., Zabeau, M., Montagu, M. Van, & Leemans, J. (1987). Transgenic plants protected from insect attack. *Nature 1987 328:6125*, *328*(6125), 33–37. https://doi.org/10.1038/328033a0

van der Loos, L. M., D'hondt, S., Willems, A., & De Clerck, O. (2021). Characterizing algal microbiomes using long-read nanopore sequencing. *Algal Research*, *59*, 102456. https://doi.org/10.1016/J.ALGAL.2021.102456

Vázquez Torres, S., Benard Valle, M., Mackessy, S. P., Menzies, S. K., Casewell, N. R., Ahmadi, S., Burlet, N. J., Muratspahić, E., Sappington, I., Overath, M. D., Rivera-de-Torre, E., Ledergerber, J., Laustsen, A. H., Boddum, K., Bera, A. K., Kang, A., Brackenbrough, E., Cardoso, I. A., Crittenden, E. P., … Baker, D. (2025). De novo designed proteins neutralize lethal snake venom toxins. *Nature*, *639*(8053), 225–231. https://doi.org/10.1038/S41586-024-08393-X

Venkata Mohan, S., Nikhil, G. N., Chiranjeevi, P., Nagendranatha Reddy, C., Rohit, M. V., Kumar, A. N., & Sarkar, O. (2016). Waste biorefinery models towards sustainable circular bioeconomy: Critical review and future perspectives. *Bioresource Technology*, *215*, 2–12. https://doi.org/10.1016/J.BIORTECH.2016.03.130

Waeterschoot, J., Gosselé, W., Lemež, Š., & Casadevall i Solvas, X. (2024). Artificial cells for in vivo biomedical applications through red blood cell biomimicry. *Nature Communications 2024 15:1*, *15*(1), 1–17. https://doi.org/10.1038/s41467-024-46732-8

Wagner, A., Wang, C., Fessler, J., DeTomaso, D., Avila-Pacheco, J., Kaminski, J., Zaghouani, S., Christian, E., Thakore, P., Schellhaass, B., Akama-Garren, E., Pierce, K., Singh, V., Ron-Harel, N., Douglas, V. P., Bod, L., Schnell, A., Puleston, D., Sobel, R. A., … Yosef, N. (2021). Metabolic modeling of single Th17 cells reveals regulators of autoimmunity. *Cell*, *184*(16), 4168-4185.e21. https://doi.org/10.1016/J.CELL.2021.05.045

Walhout, A. J. M., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., Van den Heuvel, S., & Vidal, M. (2000). GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes. *Methods in Enzymology*, *328*, 575-IN7. https://doi.org/10.1016/S0076-6879(00)28419-X

Walker, K. T., Li, I. S., Keane, J., Goosens, V. J., Song, W., Lee, K. Y., & Ellis, T. (2024). Self-pigmenting textiles grown from cellulose-producing bacteria with engineered tyrosinase expression. *Nature Biotechnology 2024*, 1–10. https://doi.org/10.1038/s41587-024-02194-3

Walsh, G., & Walsh, E. (2022). Biopharmaceutical benchmarks 2022. *Nature Biotechnology*, *40*(12), 1722–1760. https://doi.org/10.1038/S41587-022-01582-X

Wang, B., Kitney, R. I., Joly, N., & Buck, M. (2011). Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications*, *2*(1), 1–9. https://doi.org/10.1038/NCOMMS1516

Wang, B., Zhao, J., Zhang, J., Wei, T., Han, K., & Gao, T. (2024). Electrochemical biosensing interfaced with cell-free synthetic biology. *TrAC Trends in Analytical Chemistry*, *176*, 117756. https://doi.org/10.1016/J.TRAC.2024.117756

Wang, K., Fredens, J., Brunner, S. F., Kim, S. H., Chia, T., & Chin, J. W. (2016). Defining synonymous codon compression schemes by genome recoding. *Nature 2016 539:7627*, *539*(7627), 59–64. https://doi.org/10.1038/nature20124

Wang, S., Zhan, Y., Jiang, X., & Lai, Y. (2024). Engineering Microbial Consortia as Living Materials: Advances and Prospectives. *ACS Synthetic Biology*, *13*(9), 2653–2666. https://doi.org/10.1021/ACSSYNBIO.4C00313

Wang, X., Gao, Y., Zhou, N., Yang, Z., Cooper, J. M., & Wang, B. (2025). CRISPR-enabled genetic logic circuits for biosensing. *TrAC Trends in Analytical Chemistry*, *190*, 118297. https://doi.org/10.1016/J.TRAC.2025.118297

Wang, Y., Li, X., Wang, C., Gao, L., Wu, Y., Ni, X., Sun, J., & Jiang, J. (2021). Unveiling the transcriptomic complexity of Miscanthus sinensis using a combination of PacBio long read- and Illumina short read sequencing platforms. *BMC Genomics*, *22*(1), 1–16. https://doi.org/10.1186/S12864-021-07971-X

Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLOS ONE*, *6*(2), e16765. https://doi.org/10.1371/JOURNAL.PONE.0016765

Wei, S. Y., Gao, G. R., Ding, M. Z., Cao, C. Y., Hou, Z. J., Cheng, J. S., & Yuan, Y. J. (2024). An Engineered Microbial Consortium Provides Precursors for Fengycin Production by Bacillus subtilis. *Journal of Natural Products*, *87*(1), 28–37. https://doi.org/10.1021/ACS.JNATPROD.3C00705

Wei, X., Penkauskas, T., Reiner, J. E., Kennard, C., Uline, M. J., Wang, Q., Li, S., Aksimentiev, A., Robertson, J. W. F., & Liu, C. (2023). Engineering Biological Nanopore Approaches towards Protein Sequencing. *ACS Nano*, *17*(17), 16369. https://doi.org/10.1021/ACSNANO.3C05628

Wen, A., Havens, K. L., Bloch, S. E., Shah, N., Higgins, D. A., Davis-Richardson, A. G., Sharon, J., Rezaei, F., Mohiti-Asli, M., Johnson, A., Abud, G., Ane, J. M., Maeda, J., Infante, V.,

Gottlieb, S. S., Lorigan, J. G., Williams, L., Horton, A., McKellar, M., … Temme, K. (2021). Enabling Biological Nitrogen Fixation for Cereal Crops in Fertilized Fields. *ACS Synthetic Biology*, *10*(12), 3264–3277. https://doi.org/10.1021/ACSSYNBIO.1C00049

Wesseler, J., Kleter, G., Meulenbroek, M., & Purnhagen, K. P. (2023). EU regulation of genetically modified microorganisms in light of new policy developments: Possible implications for EU bioeconomy investments. *Applied Economic Perspectives and Policy*, *45*(2), 839–859. https://doi.org/10.1002/AEPP.13259

Wiegand, D. J., Rittichier, J., Meyer, E., Lee, H., Conway, N. J., Ahlstedt, D., Yurtsever, Z., Rainone, D., Kuru, E., & Church, G. M. (2025). Template-independent enzymatic synthesis of RNA oligonucleotides. *Nature Biotechnology*, *43*(5), 762–772. https://doi.org/10.1038/S41587-024-02244-W

Wong, A. Y. T., & Chan, A. W. K. (2016). Genetically modified foods in China and the United States: A primer of regulation and intellectual property protection. *Food Science and Human Wellness*, *5*(3), 124–140. https://doi.org/10.1016/J.FSHW.2016.03.002

Wu, S., Feng, J., Liu, C., Wu, H., Qiu, Z., Ge, J., Sun, S., Hong, X., Li, Y., Wang, X., Yang, A., Guo, F., & Qiao, J. (2022). Machine learning aided construction of the quorum sensing communication network for human gut microbiota. *Nature Communications 2022 13:1*, *13*(1), 1–13. https://doi.org/10.1038/s41467-022-30741-6

Xie, Z. X., Li, B. Z., Mitchell, L. A., Wu, Y., Qi, X., Jin, Z., Jia, B., Wang, X., Zeng, B. X., Liu, H. M., Wu, X. Le, Feng, Q., Zhang, W. Z., Liu, W., Ding, M. Z., Li, X., Zhao, G. R., Qiao, J. J., Cheng, J. S., … Yuan, Y. J. (2017). "Perfect" designer chromosome v and behavior of a ring derivative. *Science*, *355*(6329). https://doi.org/10.1126/SCIENCE.AAF4704

Xu, H., Wang, S., Li, K., Bai, Z., & Yang, Y. (2025). Optimization of Multiple Combinations of Early Secretion Pathways to Enhance Extracellular Protein Secretion in Komagataella phaffii. *Journal of Agricultural and Food Chemistry*, *73*(28), 17750–17758. https://doi.org/10.1021/ACS.JAFC.5C03402

Xue, T., Zheng, H., Zhao, Y., Zhao, Z., Wang, J., Zhang, Y., Li, Y., Wang, S., Liu, Y., Xue, C., & Guo, H. (2024). A spontaneously immortalized muscle stem cell line (EfMS) from

brown-marbled grouper for cell-cultured fish meat production. *Communications Biology 2024 7:1*, *7*(1), 1–15. https://doi.org/10.1038/s42003-024-07400-1

Yan, Q., & Fong, S. S. (2017). Challenges and advances for genetic engineering of non-model bacteria and uses in consolidated bioprocessing. *Frontiers in Microbiology*, *8*(OCT), 293643. https://doi.org/10.3389/FMICB.2017.02060

Yang, J. S., & Reyna-Llorens, I. (2023). Plant synthetic biology: exploring the frontiers of sustainable agriculture and fundamental plant biology. *Journal of Experimental Botany*, *74*(13), 3787–3790. https://doi.org/10.1093/JXB/ERAD220

Yang, K. K., Wu, Z., & Arnold, F. H. (2019). Machine-learning-guided directed evolution for protein engineering. *Nature Methods 2019 16:8*, *16*(8), 687–694. https://doi.org/10.1038/s41592-019-0496-6

Yarnall, M. T. N., Ioannidi, E. I., Schmitt-Ulms, C., Krajeski, R. N., Lim, J., Villiger, L., Zhou, W., Jiang, K., Garushyants, S. K., Roberts, N., Zhang, L., Vakulskas, C. A., Walker, J. A., Kadina, A. P., Zepeda, A. E., Holden, K., Ma, H., Xie, J., Gao, G., … Gootenberg, J. S. (2022). Drag-and-drop genome insertion of large sequences without double-strand DNA cleavage using CRISPR-directed integrases. *Nature Biotechnology 2022 41:4*, *41*(4), 500–512. https://doi.org/10.1038/s41587-022-01527-4

Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., Toyohara, K., Miyamoto, K., Kimura, Y., & Oda, K. (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science*, *351*(6278), 1196–1199. https://doi.org/10.1126/SCIENCE.AAD6359

Yuan, Q., Zeng, H., Daniel, T. C., Liu, Q., Yang, Y., Osikpa, E. C., Yang, Q., Peddi, A., Abramson, L. M., Zhang, B., Xu, Y., & Gao, X. (2024). Orthogonal and multiplexable genetic perturbations with an engineered prime editor and a diverse RNA array. *Nature Communications 2024 15:1*, *15*(1), 1–12. https://doi.org/10.1038/s41467-024-55134-9

Yusko, E. C., Bruhn, B. R., Eggenberger, O. M., Houghtaling, J., Rollings, R. C., Walsh, N. C., Nandivada, S., Pindrus, M., Hall, A. R., Sept, D., Li, J., Kalonia, D. S., & Mayer, M. (2017). Real-time shape approximation and fingerprinting of single proteins using a

nanopore. *Nature Nanotechnology*, *12*(4), 360–367. https://doi.org/10.1038/NNANO.2016.267

Zhang, M., Wang, Y., Chen, X., Xu, F., Ding, M., Ye, W., Kawai, Y., Toda, Y., Hayashi, Y., Suzuki, T., Zeng, H., Xiao, L., Xiao, X., Xu, J., Guo, S., Yan, F., Shen, Q., Xu, G., Kinoshita, T., & Zhu, Y. (2021). Plasma membrane H+-ATPase overexpression increases rice yield via simultaneous enhancement of nutrient uptake and photosynthesis. *Nature Communications 2021 12:1*, *12*(1), 1–12. https://doi.org/10.1038/s41467-021-20964-4

Zhang, X., Kong, D., Liu, X., Xie, H., Lou, X., & Zeng, C. (2021). Combined microbial degradation of crude oil under alkaline conditions by Acinetobacter baumannii and Talaromyces sp. *Chemosphere*, *273*, 129666. https://doi.org/10.1016/J.CHEMOSPHERE.2021.129666

Zhang, Y., Chiu, T. Y., Zhang, J. T., Wang, S. J., Wang, S. W., Liu, L. Y., Ping, Z., Wang, Y., Chen, A., Zhang, W. W., Chen, T., Wang, Y., & Shen, Y. (2021). Systematical Engineering of Synthetic Yeast for Enhanced Production of Lycopene. *Bioengineering 2021, Vol. 8, Page 14*, *8*(1), 14. https://doi.org/10.3390/BIOENGINEERING8010014

Zhao, M., Ma, J., Zhang, L., & Qi, H. (2024). Engineering strategies for enhanced heterologous protein production by Saccharomyces cerevisiae. *Microbial Cell Factories*, *23*(1), 1–16. https://doi.org/10.1186/S12934-024-02299-Z

Zhu, F., Cao, M. Y., Zhang, Q. P., Mohan, R., Schar, J., Mitchell, M., Chen, H., Liu, F., Wang, D., & Fu, Z. Q. (2024). Join the green team: Inducers of plant immunity in the plant disease sustainable control toolbox. *Journal of Advanced Research*, *57*, 15–42. https://doi.org/10.1016/J.JARE.2023.04.016

Zou, Z., Wu, S., Gerngross, D., Lozhkin, B., Chen, D., Tachibana, R., & Ward, T. R. (2024). Combining an artificial metathase with a fatty acid decarboxylase in a whole cell for cycloalkene synthesis. *Nature Synthesis 2024 3:9*, *3*(9), 1113–1123. https://doi.org/10.1038/s44160-024-00575-9

Chapter 2

Unravelling *Kluyveromyces marxianus'* nitrogen needs for improved dairy biorefineries

This chapter is under preparation for submission to the Journal of Industrial Microbiology & Biotechnology

Abstract

Cheese whey permeate is the main by-product of the dairy industry. It holds great biotechnological potential due to its high lactose content. *K. marxianus* is a yeast able to consume lactose very efficiently and it is a good candidate for revalorisation of this feedstock. However, there are important barriers for its efficient use as feedstock. In order to improve this process, it is crucial to understand *K. marxianus*' nitrogen needs in this feedstock. In this work we studied a snapshot of the yeast *K. marxianus* growing on the poor-nitrogen dairy feedstock whey permeate with and without supplementation of tryptone in aerobic condition. Bioreactor cultivations portrayed a deficit of nitrogen that prevented the yeast to efficiently consume lactose. Differential gene expression analysis of the transcriptome allowed the identification of key amino acid biosynthetic pathways, particularly methionine, arginine and lysine. Differential gene expression also showed the importance of nitrogen scavenging pathways when growing on low nitrogen media. We identified key amino acids for *K. marxianus* and the overall metabolism of this non-conventional yeast to harness the full potential of this dairy sub stream in the bioeconomy context.

Introduction

Yeasts are widely used in the food and biotechnology sectors for production of diverse products ranging from fermented beverages to bio-based chemicals. While the majority of biotechnology applications use the well-studied baker's yeast *S. cerevisiae*, many traditional fermentations rely on other yeasts (Hittinger et al., 2018). Increasingly some of these other species such as *Yarrowia lipolytica, Komagataella phaffii*, and *Ogataea polymorpha* are also being developed and applied in biotechnology. In most cases, the rationale for using alternative yeasts is because of unique metabolic features or physiological traits. *K. marxianus* is one such yeast that is expanding applications from its role in traditional dairy fermentations to multiple applications in biotechnology. This yeast shows a large industrial potential due to the ability of naturally using a wide range of carbohydrates (Lane & Morrissey, 2010), among which lactose, and its rapid growth rate and thermotolerance (Fonseca et al., 2008; Lane et al., 2011). *K. marxianus* has successfully been used for the development of a wide variety of bioprocesses for the production of bioethanol, aromas, single-cell protein or enzymes among others, using agricultural and industrial by-products (Belem & Lee, 1999; Hernández-Mendoza et al., 2024; Hu et al., 2012; Koukoumaki et al., 2024; Parekh & Margaritis, 1986; Pinheiro et al., 2003; Roy et al., 2023; Smets et al., 2025; Zhang et al., 2017; Zhang et al., 2022).

Due to its great industrial potential, the development of specific *K. marxianus* laboratory tools and protocols has gained traction in the last decade. Important milestones were the release of fully sequenced *K. marxianus* genomes and associated proteome to the chromosome level and the data-driven high-quality annotation using ribosome profiling (Fenton et al., 2022; Gómez-Márquez et al., 2022; Inokuma et al., 2016; Jeong et al., 2012; Lertwattanasakul et al., 2015). These provide a robust foundation for comparative genomics and metabolic engineering, particularly for traits such as thermotolerance and efficient lactose utilisation which are known to depend on ploidy level and genetic variants (Ortiz-Merino et al., 2018; Varela et al., 2019). Significant advances have been made in developing molecular biology tools for gene expression compatible with the widely used MoClo Yeast Tool Kit (YTK) for *S. cerevisiae* (Lee et al., 2015). By collecting functional origins of replication, selection

markers, integration sites and promoters and terminators the high-throughput screening of *K. marxianus* strains became possible (Rajkumar et al., 2019). Moreover, the development of a CRISPR/Cas9 toolkit and protocol paved the way for precise engineering of this yeast (Rajkumar & Morrissey, 2022; Varela et al., 2019).

Although, like *S. cerevisiae*, *K. marxianus* is a budding yeast, the species are separated by several hundred millions of years of evolution and the *Kluyveromyces* genus forms a subclade with *Lachancea* and *Eremothecium* (Groenewald et al., 2023; Shen et al., 2020). This evolutionary distance between *K. marxianus* and *S. cerevisiae* gives rise to significant differences between these species, from genome arrangement to regulatory systems and metabolic networks. These differences shape the characteristic features of these yeasts. One of the greatest differences is the fact that *S. cerevisiae* underwent a whole genome duplication (WGD), later demonstrated to be an interspecies hybridisation, while *K. marxianus* did not (Donzella et al., 2023; Gabaldón, 2020; Kellis et al., 2004; Wolfe, 2015). This WGD led to the specialisation of duplicated genes into new functions, which evolved into divergent genetic regulatory mechanisms and metabolic pathways (Steenwyk & Rokas, 2018). For example, one of the characteristics derived from the WGD is the improved capacity of *S. cerevisiae* to grow in anaerobiosis and perform alcoholic fermentation even under aerobic condition, also known as Crabtree effect, while *K. marxianus* shows a preference towards aerobic growth (Gabaldón, 2020; Hagman et al., 2014; Malina et al., 2021; Merico et al., 2007; Sakihama et al., 2019). The WGD can also provide some rationale for the metabolic differences such as particular aroma profile generated by each yeast during wine fermentation and other beverages (Gethins et al., 2015; Imura et al., 2020; Zhang et al., 2022). Furthermore, unlike *S. cerevisiae*, *K. marxianus* exhibits thermotolerance exceeding 45°C, driven by a more rigid cell wall, stable lipid phase transitions in the membrane, and a superior capacity to manage heat-induced oxidative stress (Kosaka et al., 2022; Mejía-Barajas et al., 2018). These features have been connected to evolutionarily young genes in *K. marxianus* (Doughty et al., 2020; Montini et al., 2022). Despite these main differences, the regulatory and functional networks of *K. marxianus* are usually inferred from *S. cerevisiae* due to a lack of experimental studies and the close phylogenetic relationship to this model yeast

(Marcišauskas et al., 2019; Schabort et al., 2016). To solve this gap, carbon fate and metabolism have recently been explored in *K. marxianus* with transcriptomic and metabolomic approaches helping to unravel *K. marxianus* gene function and regulation (Gao et al., 2024; Huang et al., 2024; Lertwattanasakul et al., 2015; Li et al., 2023; Matsuda et al., 2025; Yu et al., 2021).

In contrast, nitrogen metabolism and regulation in *K. marxianus* has not been extensively explored. A bioinformatic survey showed that, in comparison to other non-WGD yeasts, *K. marxianus* lacks typical nitrogen assimilation pathways, such as amine degradation or nitrate uptake pathways. However, it still retained genes for purine and uracil assimilation, that are also missing in *S. cerevisiae* (Linder, 2019a). Genome duplication offered the potential for diversification of transporter in *S. cerevisiae,* and this is seen with ammonium permeases: *S. cerevisiae* has the *MEP1* and *MEP3* paralogues as well as *MEP2*, whereas *K. marxianus*, possesses only *MEP2* and *MEP3* (Ai et al., 2022). The lack of paralogues for genes encoding proteins with differential location is a particular issue and, previously, we showed that differential transcriptional regulation and use of alternative transcriptional start sites enables *K. marxianus* to produce two isoforms from a single gene, *BAT1*, whereas *S. cerevisiae* requires two genes: *BAT1* and *BAT2* (Coral-Medina et al., 2022). There have been some studies comparing aspects of nitrogen metabolism in *K. marxianus* and *S. cerevisiae* in fermented beverage matrices. Zhang et al., 2022 compared *K. marxianus* and *S. cerevisiae* gene expression and metabolome profile during cider fermentation. They observed important differences in the expression of many amino acid biosynthetic pathways, in particular biosynthesis of valine, leucine, isoleucine, alanine, aspartate, glutamate, glycine, serine, threonine and lysine was higher in *K. marxianus*. They hypothesised that this might be due to a reduced intake and assimilation of amino acids from the apple juice. Rollero et al., 2019 conducted the only deep study of nitrogen metabolism in *K. marxianus* by analysing metabolic fluxes and gene expression of nitrogen pathways in this yeast during wine fermentation and comparing it to *S. cerevisiae*. This study uncovered key differences in *K. marxianus*, including a decreased ammonium uptake, an increased metabolic turnover of the amino acid arginine and up-regulation of methionine uptake and biosynthesis. These

comparative studies indicated that there was broad similarity to *S. cerevisiae* but also some regulatory and metabolic differences that needed to be explored but could not be addressed in those studies.

Nitrogen metabolism is crucial for yeast fermentations. Many of the matrices used for fermented beverages, for example, apple juice or grape must, are low in available nitrogen. Furthermore, many industrial streams, such as lignocellulosic hydrolysates or specific dairy effluents, are also poor in nitrogen (Peng et al., 2003; Tsermoula et al., 2023). This imposes severe metabolic constraints on microbial hosts, often limiting protein synthesis and inducing stress responses. While exogenous nitrogen supplementation can mitigate these bottlenecks (Rojo et al., 2023; Walker, 2014), it significantly increases operational costs and the environmental footprint (de Lima et al., 2022; Raposo et al., 2016). Consequently, a detailed understanding of nitrogen metabolism is critical for engineering robust strains capable of mobilising endogenous nitrogen pools in nutrient-limited environments. This challenge is exemplified by the use of Whey Permeate (WP), the main by-product of the dairy industry. WP is a lactose-rich protein-deficient effluent that typically contains about 0.4 to 0.8 g L^{-1} of protein and 50 to 70 g L^{-1} of lactose (Kassem, 2015; O'Donoghue & Murphy, 2023; Tsermoula et al., 2023). Understanding how *K. marxianus* orchestrates mobilisation of nitrogen in dairy substrates is crucial for the development of efficient bioprocesses.

It is evident that to understand a complex metabolic and regulatory process, carefully controlled comparative experiments must be performed. The main aim of this study was to take a systems approach to determine how *K. marxianus* responds to low nitrogen environment. By comparing to *S. cerevisiae*, the extent to which regulation is conserved can then be predicted. To increase the biotechnological relevance of the findings, the study was performed using WP as the base medium. Once it was established that this was a low-nitrogen growth medium, it was possible to supplement and then compare the transcriptome under nitrogen replete and nitrogen-poor conditions. This analysis led to the identification of conserved and unique pathways and responses and will enable the developed of improved bioprocesses in the future.

Material and methods

Strains, maintenance and preculture

K. marxianus CBS 397 was purchased from The Westerdijk Fungal Biodiversity Institute (Utrecht, The Netherlands) in lyophilized form. Subsequently, it was cultured in YPLac medium (yeast extract, 10 g L^{-1}; peptone, 20 g L^{-1}; lactose, 20 g L^{-1}) for 20 – 24 hours and stocked in sterile 60% (v/v) glycerol to a final concentration of glycerol of 24% (v/v) and 1 mL aliquots of this mixture were frozen at -80 °C. For cultivation, a frozen stock vial was used to streak a YPLac agar plate (YPLac + agar 2%) that was incubated for 24 – 48 hours at 30 °C. After that, the plate was stored at 4 °C for up to 1 month. Inoculum was prepared by transferring a single colony from the YPLac plate to 20 mL of autoclaved and decanted WP held in a 50 mL Corning tube. After 18 – 20 hours of growth on an orbital shaker 200 rpm at 30 °C, cells were directly inoculated to the desired initial concentration in the media (WP or WPT).

Preparation of Whey Permeate

WP (approximately 65 g L^{-1} of lactose) was provided by Carbery (Carbery Food Ingredients global HQ Ballineen, Co. Cork, Ireland) in 3L containers. It was stored at -20 °C upon arrival. When needed, WP was moved to 4 °C cold room and slowly defrosted. To prepare for experiments, WP was first autoclaved in 1 L Duran bottles (15 min, 121 °C). The resulting autoclaved liquid was decanted and filtered through a 0.45 µm membrane by vacuum filtration in aseptic conditions and stored up to one week at 4 °C before being used. To supplement WP with amino acids, tryptone was chosen as nitrogen supplement as it is produced by digestion of the milk protein casein and is generally available with WP. To prepare tryptone supplemented WP (WPT), tryptone (Formedium Ltd, Hunstanton, UK, Ref No.: TRP02) was added as a powder and dissolved in WP at a final concentration of 5 g L^{-1} prior to filtration and the resulting WPT was stored up to one week at 4 °C before being used. Amino acid composition of WP was analysed by Eurofins Food Testing Ireland. Total organic nitrogen (~ 0.25 g L^{-1} WP) and total inorganic nitrogen (~ 2.15 g L^{-1} WP) analyses were performed by Carbery.

Bioreactor cultivation

Bioreactor cultivations were performed in batch by triplicate using an EZcontroller bioreactor (Applikon, Delft, Netherlands). Agitation was kept at 500 rpm, temperature maintained at 42 °C, pH was left uncontrolled, air was introduced with an external air pump by applying a total flow of 1 L min^{-1} and dissolved oxygen was kept at 60% by increased stirring. The vessel capacity is 4 L, and each tank was filled up to 2 L of either WP or WPT, supplemented with 0.1 mL L^{-1} of sterile Antifoam C Emulsion (1003171934, Sigma-Aldrich, USA). Cell dry weight (CDW) was measured by sampling 10 mL from the bioreactor, spinning down 10 mL in a pre-weighed 50 mL Corning tube, removing supernatant and keeping the tube in a 60 °C incubator until biomass was completely dry.

Sugar and metabolite detection

Concentration of lactose and ethanol was determined by high-performance liquid chromatograph (HPLC). An Agilent 1200 HPLC system (Agilent Technologies, CA, USA) equipped with a refractive index detector and a REZEX 8 µL 8% H+ organic acid column (300 × 7.8 mm) (Phenomenex, CA, USA) were used to separate and detect the analytes. The column was eluted with 0.01 N H_2SO_4 at a flow rate of 0.6 mL min^{-1}. The temperature of the column was maintained at 65 °C.

Sample collection and quality control

Samples for transcriptomic analysis were collected (1 mL) in triplicate during the early exponential phase when nitrogen starts limiting growth, around 8 hours of cultivation. Samples were centrifuged (1 minute, 15000g), supernatant was discarded, and pellets were directly stored at -80 °C for subsequent extraction when required. For RNA extraction, samples were thawed and kept on ice until needed. 650 µL of Acid Phenol (Phenol-Chloroform-Isoamyl alcohol mixture, pH 4.5, ratio 125:24:1, P1944 Sigma-Aldrich) per sample were pre-warmed for 20 min at 65 °C. In the meantime, RNA lysis buffer was prepared by mixing the following amounts per RNA sample: 12.5 µL of 0.5 M EDTA, 15 µL of 3 M NaOAc pH 5.5, 720 µL of water. Cells were centrifuged, supernatant discarded, and pellet resuspended in 600 µL of RNA Lysis buffer. 80 µL of 10% SDS were added, and the samples were heated up at

65 °C for 3 min. After this, 600 µL of prewarmed acid phenol per sample were added and this mix was incubated at 65 °C for 5 min alternating cycles of vortexing and heating. Samples were then chilled for 2 min and centrifuged at 20000g room temperature (RT). Top aqueous phase was transferred to a new tube and 650 µL of Acid Phenol were added. Samples were incubated for 5 min at RT with intermittent vortexing. Subsequently, samples were centrifuged for 3 min 20000 g at RT and aqueous phase was transferred to a new tube. 650 µL of RNAse free chloroform (C2432 Sigma-Aldrich) were added and samples vortexed for 30 seconds. Samples were then centrifuged for 2 min 20000g RT and top aqueous phase was transferred to a new tube. After that, 1 volume of 2-propanol (I9516 Sigma-Aldrich) and 1.5 µL of Glycoblue Coprecipitant 15 mg mL^{-1} (AM9516 ThermoFisher Scientific) were mixed and incubated for 15 min in ice. Finally, to precipitate the RNA fraction, samples were centrifuged for 30 min, 20000g at 4 °C. Then, liquid phase was discarded, and pellet washed with 750 µL of 80% iced ethanol. After a second centrifugation step, 5 min 20000g at 4 °C, alcohol was discarded, samples completely air dried, and RNA pellet resuspended in 30 µL of Milli-Q water for quantification and storage at -80 °C to use when needed. RNA integrity (RIN) was determined through automated electrophoretic analysis on Agilent 2100 Bioanalyzer and only samples with RIN higher than 7 and 280:260 ratios higher than 2 were further used for RNA-sequencing purpose.

Transcriptome analysis

Transcriptome was sequenced as a service by Teagasc (Teagasc Food Research Centre, Moorepark, Fermoy, Ireland). Briefly, ribosomal RNA was depleted using the Ribo-Zero rRNA Removal Kit (Illumina), and cDNA libraries were constructed. Sequencing was performed on the Illumina NextSeq 2000 platform using Unique Dual Indexes (IDT for Illumina) to generate 150 bp single-end reads. Raw reads were quality-controlled and aligned to the *K. marxianus* genome (strain DMKU3-1042, (Fenton, et al., 2022)) using STAR (v2.7.10a). Gene counts were quantified simultaneously with alignment using the STAR --quantMode GeneCounts option (Dobin et al., 2013). Stranded read counts (corresponding to the reverse strand) were extracted and compiled into a count matrix using custom Python scripts. Differential

expression analysis was performed using DESeq2 (v1.34.0) in R (Love et al., 2014). Resulting data were processed using custom Python scripts; genes were filtered and visualised using the Pandas and Matplotlib libraries. Genes were considered differentially expressed with an adjusted p-value < 10^{-25} and $|\log_2 \text{FoldChange}| > 1.5$.The genome annotation is available on the open access GWIPS-viz browser (https://gwips.ucc.ie/cgi-bin/hgGateway) and the transcriptome data is available on Zenodo (https://zenodo.org/records/16745004).

Results

Batch culture of *K. marxianus* in Whey Permeate

In order to assess whether available nitrogen was the limiting growth factor for *K. marxianus* in the WP available for our study, *K. marxianus* was grown on whey permeate with (WPT) and without (WP) addition of tryptone (5 g L^{-1}) in bioreactors. (Fig. 1). Biomass was recorded as CDW and metabolites were measured over 35h of cultivation. Initially, the growth profile in both conditions was very similar, however after 8 hours of incubation, the growth rate under the supplemented (WPT) condition increased significantly and the strain continued to grow for 25h at which stage all the lactose was consumed. In contrast, under the non-supplemented condition (WP), the growth rate was slower, and biomass accumulation stopped at around 12 hours. Lactose consumption slowed and even after 35h of incubation, more than 50% of the original lactose remained. The other notable difference between the conditions related to ethanol production and consumption. Under both conditions, ethanol started to accumulate from around 8 hours, presumably due to overflow metabolism, though with substantially more produced in the supplemented conditions. Whereas the ethanol remained in the media in the WP condition, in the supplemented WPT condition, it was subsequently consumed, such that none remained at 35 hours. These data from growth in bioreactors demonstrates that the WP used in this study is low in nitrogen and thus yeast growing in this medium are in nitrogen-deficient conditions. Analysis of the amino acid composition of WP accords with this as it shows a low level of most amino acids, ranging from 10 mg to 100 mg per 100 g of WP, with the levels of aromatic amino acids (tryptophan, phenylalanine and tyrosine) below the detection limit of our method (Fig. S1).

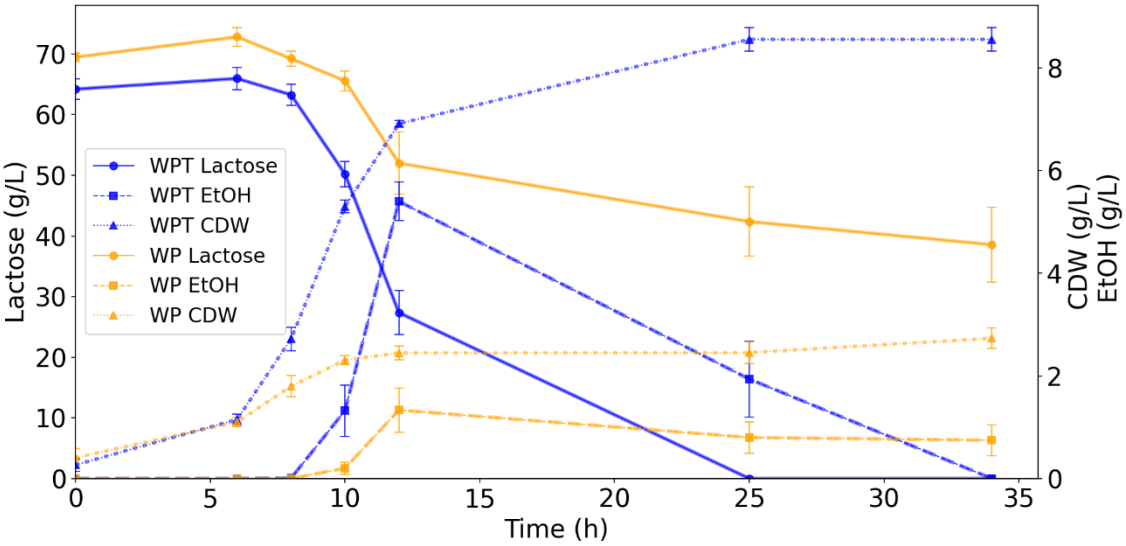


Fig. 1. Bioreactor growth curve of *K. marxianus* growing on WPT and WP. Growth curve of *K. marxianus* on WPT (blue) and WP (orange). Solid lines with • represent lactose (g L^{-1}) and is plotted on left y axis. Dashed lines with ▪ represent ethanol (g L^{-1}) and is plotted on right y axis. Dotted lines with represent CDW (g L^{-1}) and is plotted on right y axis.

Differential gene expression

To investigate the genetic response of *K. marxianus* to N-limited (WP) or N-replete (WPT) conditions, transcriptome analysis was performed on exponentially-growing strains from each condition and differential gene expression (DGE) analysis was performed applying a stringent significance threshold ($\log_2$ FC > 1.5; p < 10^{25}) (Table S1). This analysis revealed that, under N-limiting conditions, 90 genes were differentially expressed; expression of 79 genes was higher, and 11 genes lower (Fig. 2, Table 1). Using Gene Ontology (GO) terms analysis, we categorised the differentially-expressed genes to generate an overall view of the cell's response to the different nitrogen status. We also categorised and grouped the genes by metabolic pathway to compare the level of transcripts within each relevant module to give an indication of the relative use of each of the routes under these conditions (Table S2).

Gene ID	Gene Name	Log_2 FC	padj
Nitrogen uptake			
KLMX_20363	MUP1	2.97	5.80E-25
KLMX_70370	PTR2	-7.29	1.85E-167
KLMX_70368	PTR2	-6.44	3.31E-68
KLMX_70369	PTR2	-5.41	1.21E-35
KLMX_80174	AQR1	1.73	1.29E-28
KLMX_20760	MEP2	8.40	2.84E-154
KLMX_30296	DUR3	3.70	1.57E-49
KLMX_50327	SEO1	2.16	2.45E-56
KLMX_10383	GAP1	5.42	3.61E-32
KLMX_20006	OPT1	2.43	9.36E-55
Protein catabolism			
KLMX_50345	APE2	-2.05	7.50E-78
KLMX_50068	BNA3	1.97	2.07E-32
KLMX_10369	CPS1	2.03	9.16E-27
KLMX_10450	DAO1	1.87	1.89E-33
KLMX_50085	DUG2	1.71	5.61E-30
KLMX_40303	YGL159W	2.38	8.19E-27
Lysine biosynthesis			
KLMX_10781	LYS21	3.43	2.39E-55
KLMX_60497	LYS21-2	6.58	4.79E-130
KLMX_30516	LYS4	4.00	1.78E-88
KLMX_80129	LYS12	4.54	1.25E-86
KLMX_80299	LYS2	3.84	3.09E-65
KLMX_40016	LYS9	5.18	2.98E-123
KLMX_40377	LYS1	4.73	5.34E-183
Arginine biosynthesis			
KLMX_20813	ARG2	1.83	1.26E-33
KLMX_10446	ARG4	1.94	7.94E-28
KLMX_60190	ARG7	2.55	1.10E-35
KLMX_80390	ARG8	2.45	1.57E-49
KLMX_50176	CAR2	2.45	7.54E-35
Cysteine and Methionine biosynthesis			
KLMX_40351	HOM3	1.99	2.03E-43
KLMX_50077	cysB	1.87	1.01E-34
KLMX_30315	STR3-2	5.06	4.34E-119
KLMX_20363	MUP1	2.97	5.80E-25
KLMX_80347	MET6	3.81	3.48E-169
KLMX_20275	MET13	4.13	9.88E-70
KLMX_10684	MET17	6.82	1.00E-300

Gene ID	Gene Name	Log_2 FC	padj
Histidine biosynthesis			
KLMX_80165	HIS5	1.91	1.52E-31
KLMX_20410	HIS7	1.94	1.14E-41
Aromatic amino acid metabolism			
KLMX_40111	ARO3	2.30	8.21E-41
KLMX_40637	ARO5*	1.96	6.00E-27
KLMX_20590	ARO10	-4.56	1.76E-92
Glutamate biosynthesis			
KLMX_50018	GDH3	6.76	3.13E-149
Branched chain amino acids			
KLMX_30108	ILV2	2.31	2.12E-54
KLMX_40267	ILV3	2.62	5.69E-43
KLMX_30621	ILV5	3.01	6.72E-37
KLMX_30146	ILV6	2.33	2.27E-36
KLMX_50152	LEU1	4.02	2.02E-116
KLMX_20479	LEU2	3.94	7.23E-103
KLMX_50399	LEU4-1	2.61	2.01E-60
KLMX_60229	LEU4-2	2.02	5.73E-49
Sulphur uptake			
KLMX_30247	MET1	3.29	5.18E-142
KLMX_10088	MET2	6.08	4.66E-177
KLMX_80062	MET3	3.97	5.80E-219
KLMX_40505	MET5	4.44	3.95E-110
KLMX_50465	MET10	4.55	3.17E-116
KLMX_30074	MET14	6.66	3.88E-221
KLMX_70265	MET16	4.29	2.93E-69
KLMX_60280	MET22	1.81	1.34E-47
KLMX_60391	MET32	2.02	3.62E-37
KLMX_40153	SUL2	4.09	4.08E-117
Trans-sulphuration			
KLMX_80401	YHR112C*	2.20	7.65E-47
Transcription factor			
KLMX_40008	ZCF1*	4.20	9.36E-74
KLMX_60300	GCR2-like*	3.34	6.68E-48
KLMX_70318	MET28*	4.19	3.48E-25
KLMX_30014	ZCF32-like*	4.24	8.47E-39
KLMX_50443	DAL81-like*	2.30	1.71E-32
Carbohydrate metabolism			
KLMX_10722	SOU2	3.24	3.25E-30
KLMX_10151	ACO2a	3.44	6.18E-79

Gene ID	Gene Name	Log_2 FC	padj
Purine catabolism			
KLMX_20060	URO1*	2.74	7.37E-33
Cell adhesion			
KLMX_70437	FLO-like*	-2.05	6.43E-54
Cell division			
KLMX_70362	PCL5	3.63	2.47E-103
Cell wall organisation			
KLMX_10836	CWP-like*	-6.70	8.97E-60
KLMX_50200	HRQ2*	4.42	2.63E-116
Glutathione metabolism			
KLMX_80015	hyuA	-2.83	8.91E-36
Monooxygenase			
KLMX_60403	PDAT9	2.84	1.45E-34
Oxidoreductase			
KLMX_60522	FRE1	-3.49	1.20E-36
Tetrahydrofolate cycle			
KLMX_10520	FAU1	3.68	1.04E-54
Vacuolar sorting			
KLMX_50491	VPS70	1.37	6.33E-27
Glyoxylate related			
KLMX_20303	AGX1	3.40	2.39E-46
KLMX_40480	IDP1	2.69	2.50E-43
Metal ion transport			
KLMX_70298	CTR1	-3.36	1.69E-38
KLMX_10610	YDR061W	2.06	2.14E-28
Mitochondrial transport			
KLMX_60341	CTP1	3.43	5.01E-60
KLMX_80069	mcfL	2.36	2.03E-34
Meiosis			
KLMX_10167	SPO21*	1.79	9.15E-29
KLMX_30322	SPO23*	3.24	6.59E-26
Lipid metabolism			
KLMX_40107	CHO2	-2.43	5.09E-34
KLMX_30499	LEC1	1.61	1.16E-33
Oxidative stress			
KLMX_50415	CTT1	3.32	4.90E-31
KLMX_20196	ppr1/MPR1	3.99	1.42E-86
Unknown			
KLMX_40310	YKL068W-A*	2.61	1.56E-30

Table 1. List of significantly differentially expressed genes (adjusted p-value (padj) < 1×10^{-25}, absolute log_2 fold change (Log_2 FC) ≥ 1.5) in *K. marxianus* during growth on WP versus WPT. Positive Log_2 FC values indicate higher expression in WP, while negative values indicate higher expression in WPT. Biological process was assigned based on annotation from UniProt and was manually curated using the GWIPS-viz browser, synteny tools, AlphaFold and SeekFold. * Unannotated genes.

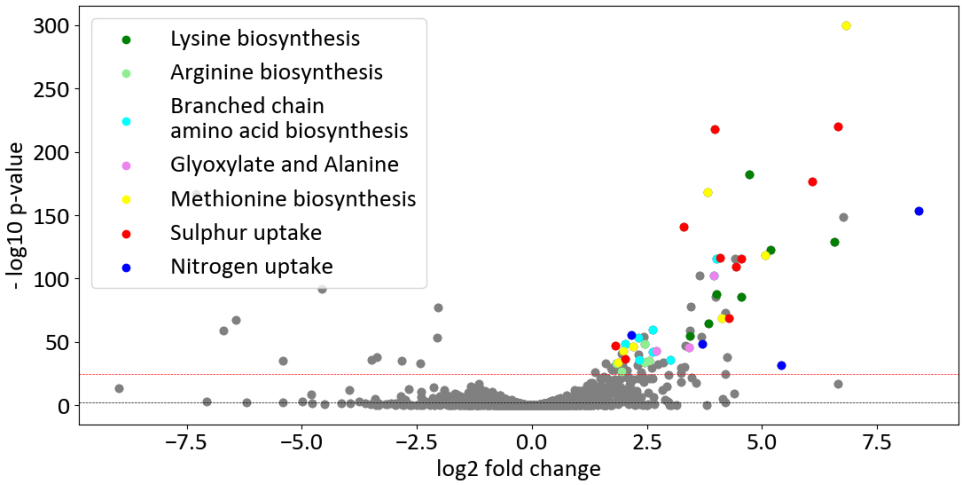


Fig. 2. Volcano plot of differentially expressed genes in WP against WPT. The x-axis represents the $\log_2$ fold change in gene expression between conditions. The y-axis shows the $-\log_{10}$ (adjusted p-value) from differential gene expression analysis. Genes with significant upregulation during cultivation in WP ($\log_2$ FC > 1.5, $p < 1{\cdot}10^{-25}$) are located on the right side of the plot ($\log_2$ FC < -1.5, $p < 1{\cdot}10^{-25}$), while significantly downregulated genes are located on the left side. Colours in the legend highlight specific GO terms identified.

As expected, GO terms align with the nitrogen status; under N-limiting conditions most of the genes showing higher levels of expression encode proteins involved in nitrogen uptake or amino acid metabolism. Moreover, several genes encoding enzymes involved in the generation of glyoxylate, most significantly *AGX1*, were differentially expressed, as well as two oxidoreductases (*IDP1* and *SOU2*) necessary for building up redox power needed in anabolic pathways. We were also able to identify in our reference genome a group of 12 unannotated genes that were also differentially expressed, and their function was identified through bioinformatic methods (Table 1, Table S3). Among these, a gene similar to *HRQ2* from *Candida albicans*, 2 meiotic-related genes *SPO21* (*KLMX_10167*) and *SPO23* (*KLMX_30322*), and a gene of unknown function *YKL068W-A* encoding a short peptide.

In N-replete conditions, three homologues of the *PTR2* oligopeptide transporter (*KLMX_70368*, *KLMX_70369* and *KLMX_70370*) are highly expressed, suggesting the

presence of oligopeptides in tryptone. The higher expression of *APE2*, encoding an amino peptidase, and *ARO10*, encoding a decarboxylase involved in the degradation of aromatic amino acids, is also consistent with the premise that WPT is a good source of organic nitrogen. Two genes, *FRE1* and *CTR1*, predicted to encode proteins involved in copper homeostasis; *CHO2*, predicted to encode an enzyme involved in phosphatidylcholine biosynthesis; and *hyuA* (*KLMX_80015*) a hydantoinase involved in recycling of glutathione were also highly expressed. The last two significantly expressed genes present a secretion signal and encode for putative surface proteins (Table S3): *KLMX_70437,* possibly encoding a Flo9-like flocculin, and *KLMX_10836* a cell wall structural protein of the PIR (Proteins with Internal Repeats) family.

Nitrogen uptake and mobilisation under nitrogen deprivation

Our data revealed that under nitrogen scarcity, expression of at least four different transport systems is activated (Table 1). In fact, *MEP2*, encoding a well-studied ammonium transporter is the most differentially expressed gene in our dataset, confirming the importance of ammonium as a nitrogen source. Like *S. cerevisiae, K. marxianus* carries a second ammonium transport gene, *MEP3*, also strongly expressed but falling below the stringent statistical cut-offs set in this study. Genes encoding four other nitrogen-scavenging permeases are differentially expressed: one of the two copies of *GAP1*, one of the three copies of *DUR3*, one of the two copies of *MUP1,* and *SEO1* (Table 1). These genes encode a general amino acid transporter, a urea transporter, a methionine transporter and a possible specific γ-Glu-Met dipeptide transporter, respectively (Berg et al., 2023; Dubey et al., 2025). Three amino acid permeases are strongly expressed under both conditions: the lysine permease *LYP1*, the arginine permease *CAN1*, and the other *GAP1* copy (Table S2). The very high level of expression of *LYP1* must reflect the importance of acquiring lysine by *K. marxianus*.

We also wanted to understand how intracellular physiology was affected by nitrogen scarcity and first explored whether, as in *S. cerevisiae*, vacuolar stores of arginine could be important for growth under low nitrogen. In fact, the expression of one of the key players for vacuolar arginine mobilisation, *VSB1* (*KLMX_20502*), is also increased in our N-limited condition WP (Tables S1, S2). Although expression of *CAR1*,

encoding the arginase that initiates degradation of arginine is equally well-expressed under both conditions, *CAR2*, encoding the next step is differentially highly expressed, as are the genes of all the other enzymes of this metabolic process (Fig 3a). Expression of *DUR1,2*, encoding the enzymatic machinery that generates ammonium from urea in two steps, is also strongly differentially expressed (Fig 3b). This ammonium can be incorporated into glutamate by two systems in *S. cerevisiae* and both also are active here (Fig 3c). *GDH3*, encoding a NADP-dependent glutamate dehydrogenase that forms glutamate from 2-oxoglutarate and ammonium is one of the most differentially expressed genes, illustrating the importance of this mechanism under N-starvation. Interestingly, there is a second gene, *GHD2*, also predicted to encode a NADP-dependent glutamate in *K. marxianus* but this is not differentially expressed. The expression of *GLN1* and *GLT1*, encoding the second pathway for glutamate synthesis, is higher in WP, but high in both conditions and so the difference is not statistically significant using our stringent thresholds (Table S1, S2). Taking these data together, it is concluded that, under these low nitrogen conditions, *K. marxianus* activates vacuolar stores of arginine and increases glutamate production to supply other amino acid biosynthetic pathways.

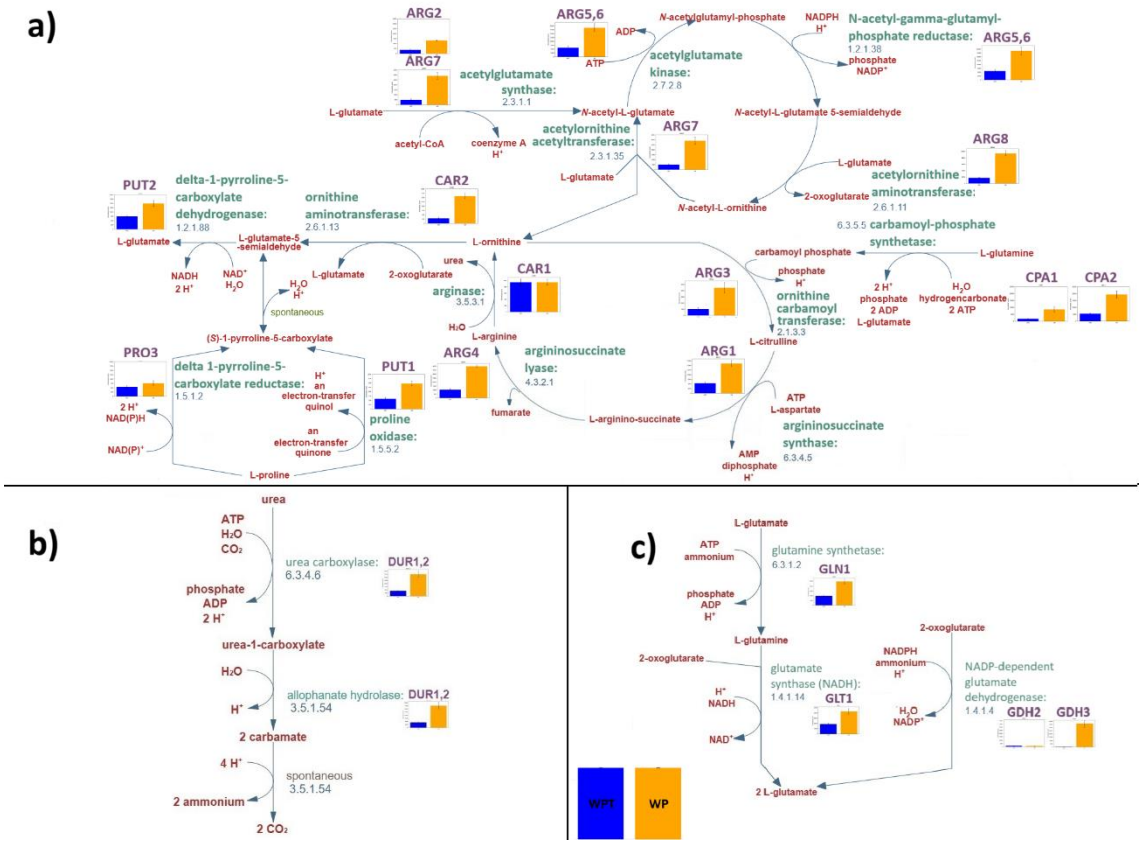


Fig. 3. Pathways for mobilisation of nitrogen vacuolar stores in *K. marxianus*. a) Arginine biosynthetic pathway and urea cycle in *K. marxianus*. **b)** Urea degradation in *K. marxianus*. **c)** Ammonium upcycling by glutamine and glutamate reactions. Each enzymatic step is annotated with a corresponding histogram representing the expression level of every enzyme under two different conditions, WPT in blue, WP in orange. When more than one enzyme is known to catalyse the same or the reverse step, their level of expression can be compared visually. The histograms provide a comparative visualisation of transcriptional regulation and metabolic fluxes across the pathway.

Biosynthesis of key amino acids under nitrogen deprivation

Many of the pathways identified by the GO term categorisation of differentially expressed genes (Table 1, Table S2) are involved in amino acid biosynthesis and these were examined in more detail. Almost all the genes of the lysine and branched chain amino acid (BCAA) pathways are expressed at significantly higher levels under low nitrogen. In the case of BCAA biosynthesis (Fig. 4a), all the genes, except *BAT1* and *ILV2,* are significantly upregulated. Likewise, the lysine biosynthetic pathway (Fig. 4b) is one of the most highly expressed pathways when growing on WP. Indeed, with the

exception of *ARO8*, all the genes that encode enzymes of this pathway are at least 10-fold more expressed when growing on WP compared to WPT.

Interestingly, although aromatic amino acids were not detected in WP, only the biosynthetic gene *ARO3* was significantly upregulated. It is known that Aro3p is a key regulatory enzyme as its activity is subject to feedback regulation, and our data suggest that controlling the level of the protein via transcriptional regulation of *ARO3* is also important. *ARO5*, recently shown to have a regulatory role in *S. cerevisiae*, is also differentially expressed (Table 1) and, when looking at the entire pathway (Fig S2, Table S2), it is seen that expression of almost all genes of the pathways show an upward trend in WP relative to WPT. Therefore, it can be inferred that synthesis of phenylalanine, tyrosine and tryptophan is also increased under low nitrogen conditions.

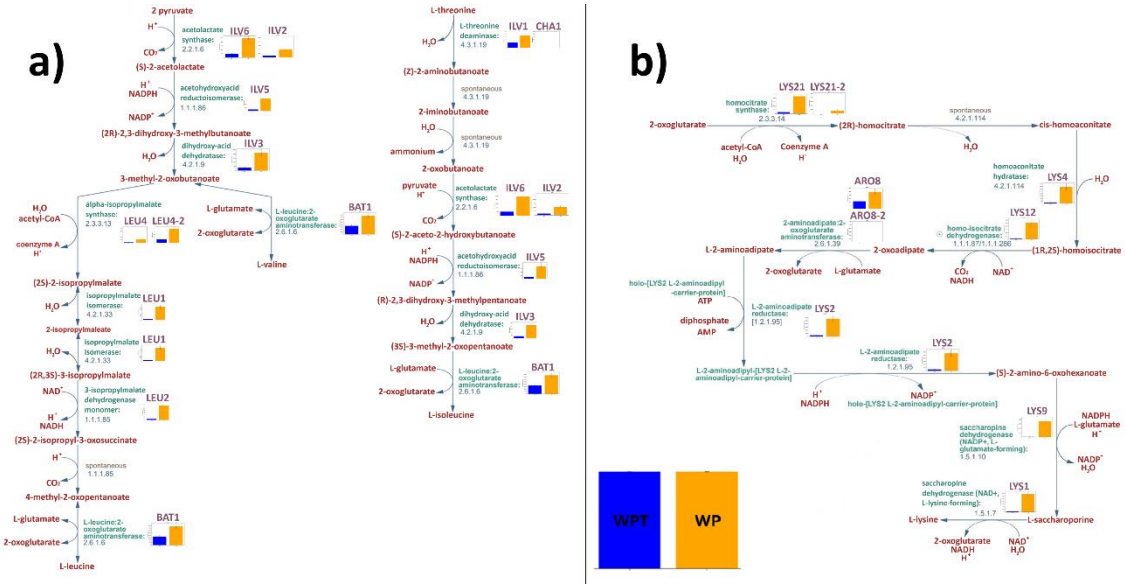


Fig. 4. Branched chain amino acids and Lysine biosynthetic pathways in *K. marxianus*. a) Leucine, valine and isoleucine (BCAA) biosynthetic pathways in *K. marxianus*. **b)** Lysine biosynthetic pathway in *K. marxianus*. Refer to Fig. 3 for explanation of bar graphs.

Sulphur uptake and sulphur-containing amino acids during nitrogen starvation

Sulphur uptake is another pathway that shows all of its genes involved highly expressed in nitrogen-deprived conditions, with an average increase 15-fold compared to WPT (Fig. 2, Table 1). While in *S. cerevisiae*, *SUL1* and *SUL2*, the two main sulphur transporters, maintain the concentration of this microelement, only *SUL2* is present in *K. marxianus*. Sul2p mediates sulphur uptake, channelling sulphur towards amino acid production by using reductive power to form hydrogen sulphide (H_2S) (Fig. 5). *SUL2* shows an increase more than 16-fold in expression during growth in WP (Table 1). There are other putative sulphur transporters in yeast: by doing a homology search against *SUL2* in *K. marxianus* we identified *KLMX_70422*, a gene present across many Saccharomycetaceae species, showing increased expression although below our significance threshold (Table S1, S2).

The sulphur pathway is tightly connected with methionine and cysteine biosynthesis. Once H_2S is synthesised, it is then incorporated into homocysteine, which is an important intermediate in the biosynthesis of cysteine and methionine. We observed that most of the genes expressing enzymes and positive regulators of this biosynthetic pathway are also highly expressed under nitrogen-poor conditions (Fig. 2, Table 1). Interestingly, from homocysteine the gene expression of *K. marxianus* in low nitrogen condition seems tilted towards the production of methionine rather than cysteine (Fig. 5). Typically, cysteine is synthesised via cystathionine. However, we also observed the increased expression of *cysB*, a cysteine synthase (EC 2.5.1.47) found in many fungi (Brzywczy et al., 2007; Les et al., 2025). Structural and sequence analysis using Foldseek and protein BLAST confirmed putative identification. *K. marxianus* already encodes a cysteine synthase also present in *S. cerevisiae*, *KLMX_40046*, paralog of *cysB* but its expression did not change significantly (Table S1, S2). *In silico* analysis of sub-cellular localisation with DeepLoc point *cysB* to be a cytoplasmic enzyme, while *KLMX_40046* being located in the mitochondrion.

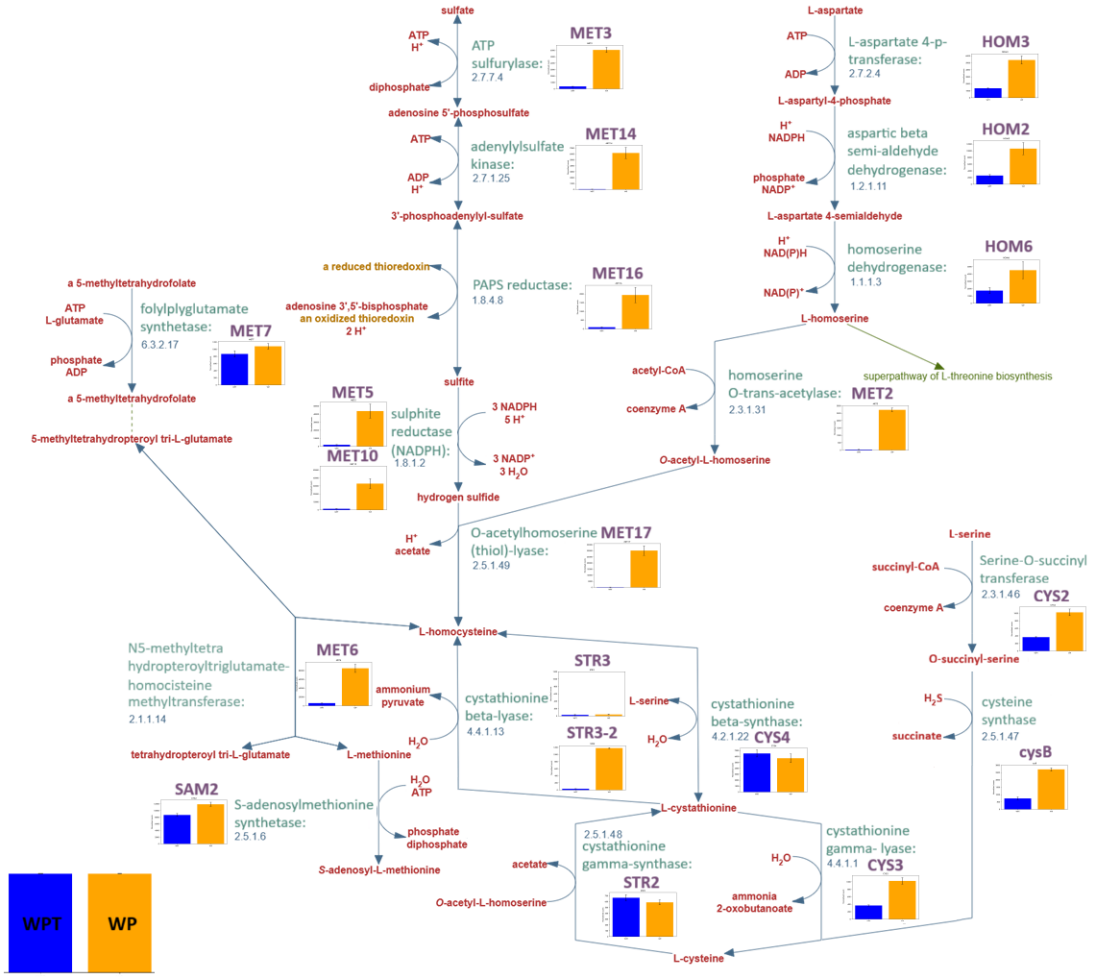


Fig. 5. Methionine and cysteine biosynthetic pathway in *K. marxianus*. Methionine, cysteine and other intermediaries in *K. marxianus*. Refer to Fig. 3 for explanation of bar graphs.

Additional routes for nitrogen scavenging in *K. marxianus*

We also found evidence of activation of other pathways that are not described in *S. cerevisiae* but that have the potential to generate ammonium from intracellular resources when nitrogen is low. Three copies of *DAO1*, encoding a putative amino acid oxidase, are present in *K. marxianus*, with one of them, *KLMX_10450,* differentially expressed. Thus, like some other yeasts, *K. marxianus* may use a FAD-dependent oxidase to hydrolyse amino acids (Pollegioni et al., 2002). *K. marxianus* also has a purine degradation pathway that is expressed at a higher level under N-limitation (Fig. 6). Through this catabolic pathway, adenine and guanine are converted into xanthine which is ultimately metabolised to release urea. While *S.*

cerevisiae has the DAL genes of the lower part of this pathway (allantoin to urea), it lacks the genes encoding the nucleotide to xanthine (*GUD1, AAH1, XAN1*); xanthine to urate (*XAN2*); and urate to allantoin (*URO1, URO2, URO3*) steps. With the exception of *XAN1*, all genes are present in *K. marxianus*, with genes for all the steps from xanthine to urea more expressed in N-limiting conditions. It should be noted, however, that although the change in expression is significant, the overall number of transcripts for the four genes encoding the xanthine to allantoin part of the pathway is low so this may only play a minor role in *K. marxianus* under these conditions (Table S2).

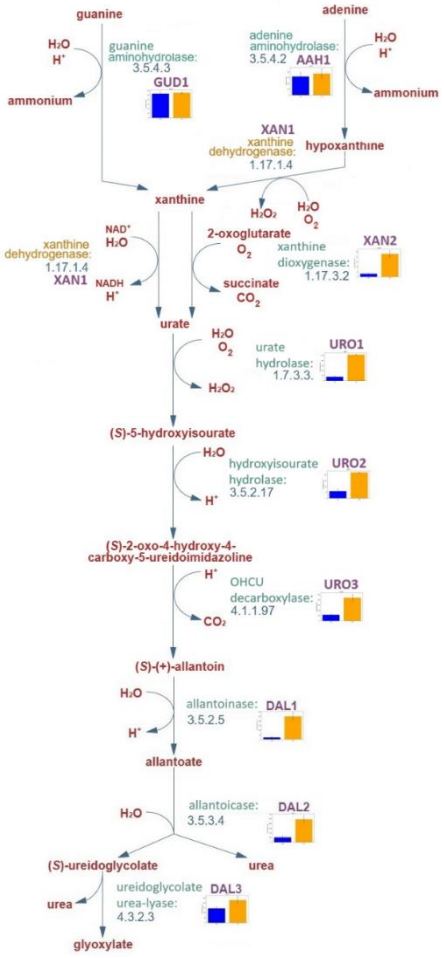


Fig. 6. Nitrogen scavenging via purine degradation in *K. marxianus*. Refer to Fig. 3 for explanation of bar graphs. There is no putative xanthine dehydrogenase found in *K. marxianus* and for this reason the enzyme's name appears in gold colour.

Lactose uptake in *K. marxianus*

Although not the primary focus of this study, we took advantage of the transcriptome data to address some outstanding issues related to lactose metabolism, which is not expected to differ between the conditions. It is known that *K. marxianus* carries four copies of *LAC12*, encoding a putative lactose transporter. In both conditions only *LAC12* is expressed, supporting previous studies that identify this as the key transporter for lactose uptake (Fig. 7a). *LAC4* encoding a β-galactosidase is expressed, therefore allowing the hydrolysis of lactose to galactose and glucose. Genes for all steps of the Leloir and glycolytic pathways are expressed and there are no anomalies. In some cases, reactions can be catalysed by isozymes, and it was possible to determine which was active at those steps when *K. marxianus* is growing in WP by the presence or absence of expression (Fig. 7b).

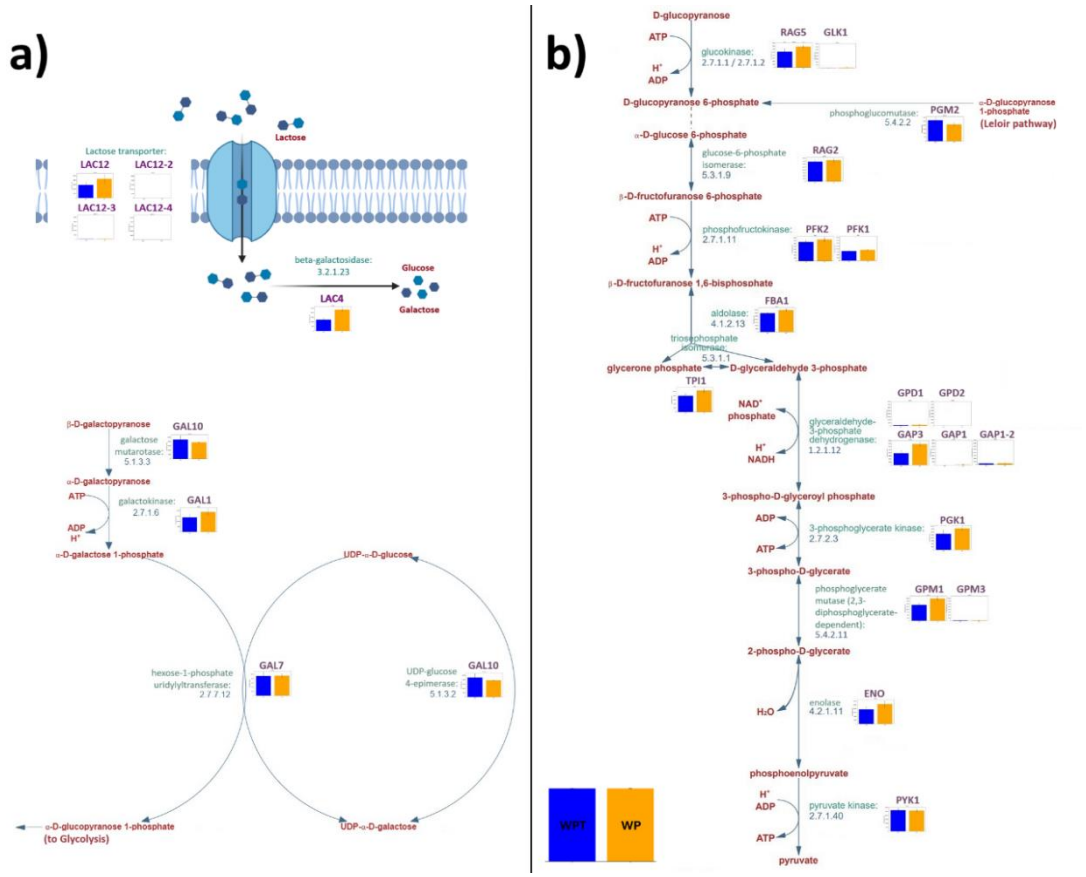


Fig. 7. Lactose uptake and glycolysis in *K. marxianus*. a. Lactose uptake and breakdown in *K. marxianus* through Leloir pathway. **b.** Glycolysis in *K. marxianus*. Refer to Fig. 3 for explanation of bar graphs.

Discussion

Cheese WP is the main by-product of the dairy industry. It holds great biotechnological potential due to its high lactose content. However, an important barrier for its use as feedstock is its low nitrogen content. *K. marxianus* is able to efficiently grow on dairy substrates. It is a yeast closely related to *S. cerevisiae,* yet their metabolic network of nitrogen utilisation and its regulation are different (Rollero et al., 2019). To fully exploit the potential of WP and similar dairy sub streams as feedstocks for *K. marxianus*, it is important to understand how this yeast mobilises nitrogen during N-limitation. In this work we studied the gene expression profile of *K. marxianus* when it was grown in bioreactor using WP as feedstock with and without addition of tryptone.

WP is poor in nitrogen and, after comparing the growth of *K. marxianus* in WP and WPT, we confirmed *K. marxianus* activates nitrogen starvation mechanisms (Fig. 2, Table 1). It is known that yeasts show different nitrogen preferences depending on the species (Kemsawasd et al., 2015; Parente et al., 2018; Roca-mesa et al., 2020), the environmental conditions (Gethins et al., 2015; Martínez-Force & Benitez, 1995) and even among strains (Fairbairn et al., 2017). For this reason, it is crucial to understand the specific nitrogen needs in order to optimise the efficacy of a bioprocess.

Under nitrogen-scarcity conditions, several amino acid transporter genes are strongly expressed during exponential growth. Specifically, *LYP1*, *CAN1*, and *GAP1* show the highest transcript abundance (Table 1). However, it is *MEP2* and *SEO1* that exhibit the most significant fold-change induction when cells are grown on WP. Notably, while previous studies reported that *K. marxianus* possesses only a single ammonium transporter, *MEP3* (Rollero et al., 2019), we confirmed the presence and the expression of *MEP2* in the strain used in this work (Table 1, Table S2). Our analysis of the genomes of other strains of *K. marxianus* shows that most strains carry both *MEP2* and *MEP3,* but a subset have lost *MEP2* (data not shown). Other nitrogen transporter genes that are also highly differentially expressed, such as *GAP1*, *DUR3*, *MUP1* present more than one copy, contrary to what is observed in *S. cerevisiae*. The reasons for this remain to be determined but it may be that in certain environments,

other nitrogen sources are plentiful, and ammonium acquisition is less critical, allowing strains to lose the gene.

Our data support that in *K. marxianus*, as in *S. cerevisiae*, vacuolar stores of arginine are mobilised when nitrogen is scarce. We observed that many genes involved in arginine metabolism were highly expressed under nitrogen scarcity in accordance with this idea. Moreover, *CAR1*, encoding the only known arginase in *K. marxianus*, was not differentially expressed, whereas the expression of the cationic amino acid vacuolar transporters *VSB1* and *YPQ1* increased (Table S1, S2), consistent with arginine intracellular mobilisation. We also investigated the existence of an alternative arginine biosynthetic pathway. Previous work identified a gamma-aminobutyrate (GABA)-dependent pathway in the sister yeast *K. lactis* (Romagnoli et al., 2014); however, the orthologous gene in *K. marxianus*, *GBU1* (*KLMX_40368*), was not upregulated in our study (Table S1, S2). These observations, combined with the constitutive expression of the arginine permease *CAN1*, support the hypothesis that arginine mobilises nitrogen reserves rather than serving solely as an anabolic building block (Rollero et al., 2019).

Nitrogen mobilisation from arginine relies heavily on the urea cycle. We observed increased expression of *DUR1,2* and urea transporters like *DUR3* (Table 1, Table S2). Since *CAR1* expression remained stable, the upregulation of *DUR3* and *DUR1,2* suggests that scavenging and processing urea derived from catabolism is the primary mechanism for ammonium generation. The ammonium generated must be assimilated efficiently. While *GLN1* and *GLT1* play the primary role in nitrogen-replete conditions, we observed a dramatic shift during starvation: *GDH3* expression surged from negligible levels to match those of *GLN1* and *GLT1*. This indicates that the NADP-dependent glutamate dehydrogenase encoded by *GDH3* becomes crucial during starvation, likely converting the ammonium mobilised into glutamate (Fig. 3). In contrast, the role of *GDH2* remains not studied in *K. marxianus*. Collectively, the interplay between arginine metabolism, the urea cycle, and transaminases allows *K. marxianus* to scavenge nitrogen from various metabolites, funnelling diverse nitrogen groups into urea and ultimately ammonium for biosynthesis.

In low nitrogen conditions, amino acid pathways are differentially highly expressed. However, *K. marxianus* shows particular preference towards the biosynthesis of lysine, branched chain amino acids and methionine (Fig. 2, Table 1). Interestingly, the content in aromatic amino acids in WP is extremely low (Fig. S1) but, with the exception of *ARO3*, this pathway is not differentially expressed. The explanation for this appears to be that because the need to be able to synthesise aromatic amino acids is high, the biosynthetic pathway is constitutively expressed and only *ARO3* needs to be differentially expressed to ramp up production. This enables a rapid response when needed.

Methionine is known to be a key molecule for fungi. It is the first amino acid added during protein synthesis, its biosynthesis is connected to the folate cycle and S-adenosyl methionine (SAM) (Fig. 4, Fig. S2), and it is a cellular sensor for anabolic capacity (Walvekar et al., 2018). In previous studies in *K. marxianus* it was found that methionine biosynthetic genes were upregulated when growing with different nitrogen sources compared to *S. cerevisiae* (Rollero et al., 2019). Likewise, many genes involved in sulphur uptake and methionine are highly upregulated in WP (Fig. 4, Table 1, Table S1). While cysteine possesses a reactive thiol group, the sulphur atom in methionine is protected by a methyl group, making it redox insensitive. This feature, combined with the high redox cost of methionine synthesis (Fig. 5) and its intrinsic link to SAM and folate metabolism (Fig. S3), highlights the critical role of methionine in cellular homeostasis, particularly in the context of nitrogen deficiency (Walvekar & Laxman, 2019).

In this work, we identified relevant genetic differences related to the nitrogen-depravation response between *K. marxianus* and *S. cerevisiae*. We detected and studied the genes from the purine catabolic pathway (Fig. 6), a metabolic route lost in the related yeast *S. cerevisiae*, previously described in many yeasts (Linder, 2019b). This metabolic pathway is transcriptionally active in *K. marxianus*, but the key step performed by *XAN2* (*KLMX_20517*) shows very low transcript levels (Table S2). The taxonomic distribution of this pathway in budding yeasts reflects its sequential nature: all species that contain either *XAN1* or *XAN2* also contain the following enzymes of the pathway. While this is the case for *Kluyveromyces*, the WGD clade

and the closely related *Eremothecium* lack the XAN and URO genes (Linder, 2019b). We also found that *K. marxianus* activates cytoplasmic production of cysteine when nitrogen is low through *CYS2* and *cysB*, a gene absent in *S. cerevisiae* but present in many other yeasts such as *Lachancea*, *Torulaspora*, *Scheffersomyces* or *Lypomyces*. We also identified a set of transcriptional factors that are not present in *S. cerevisiae*: *KLMX_30014*, *KLMX_40008*, *KLMX_50443, KLMX_60300*. These show high expression in *K. marxianus* under poor-nitrogen condition. Taken together, these findings indicate that *K. marxianus* has retained a more ancestral and diverse nitrogen assimilation repertoire, contrasting with *S. cerevisiae*, which appears to have undergone significant metabolic streamlining and gene loss in its adaptation to nutrient-rich fermentative niches.

As previously reported by our group the genome of *K. marxianus* contains 4 copies of the lactose transporter gene (*LAC12*, *LAC12*-2, *LAC12*-3 and *LAC12*-4). However, out of this four, only one can transport lactose efficiently and is used by the yeast when growing on lactose (Ortiz-Merino et al., 2018; Varela et al., 2019). We could confirm that, when growing on WP and WPT, *LAC12* is the main lactose transporter since its transcription level is fifty-fold greater than the rest of lactose transporter genes (Fig. 3, Table S1). After lactose is transported inside the yeast cell, it is broken down into glucose and galactose by Lac4p. Glucose is then metabolised through glycolysis while galactose requires to be processed through the Leloir pathway prior to its breakdown. The first step of glycolysis is glucose phosphorylation. In *S. cerevisiae* this step can be catalysed by three described hexokinases (Moreno & Herrero, 2002) but in *K. marxianus* only two are known, *GLK1* and *RAG5* (Nurcholis et al., 2019). In this work we reported that, when growing on lactose in aerobic condition, *RAG5* expression is greater than *GLK1*, and codes for the main glucokinase. Similarly, three glyceraldehyde-3-phosphate dehydrogenase genes are present in *K. marxianus* but only one is the preferred, *GAP3*, being expressed about a hundred-fold greater than the other paralogs. Finally, the phosphoglycerate mutase gene *GPM1* is preferred over *GPM3*, giving a final shape to the main expressed glycolysis genes in this yeast. In all the instances where two or more enzymes are known to catalyse the same step, one of them is especially enhanced while the rest remain almost silent.

The strict stringency that was applied to this study (p-value < 10^{-25}) might be a limitation that prevented to see a different picture than the described earlier. For this reason, a p-value < 10^{-20} was put to the test as well. Under this laxer scenario, out of the 19 highly expressed genes added up to the initially highly expressed (Table 1), 13 genes support nitrogen scavenging (*DUR1,2*, *isp4*, a second copy of *DAO1*, *DAL2* and *QDR3*), mobilisation (*ARG1*, *VSB1*, *CPA1* and *CPA2*) and amino acid biosynthesis (*HOM2*, *HIS4*, *THR4* and *MET8*) appear upregulated (Table S1). These results strengthen the conclusions extracted earlier.

Conclusion

This study elucidates the transcriptional strategies that enable *K. marxianus* to thrive in nitrogen-poor dairy waste streams. Our results show that *K. marxianus* responds to low nitrogen through mobilisation of intracellular nitrogen reserves via arginine and the urea cycle, and the upregulation of high-affinity transporters and methionine biosynthesis to maintain cellular homeostasis. Moreover, we highlighted significant divergences from the model yeast *S. cerevisiae*, identifying the retention of ancestral traits such as the purine catabolic pathway, the cytoplasmic cysteine synthase *cysB*, or a distinct set of transcription factors. These features endow *K. marxianus* with a metabolic plasticity that *S. cerevisiae* has lost. Our findings underscore the potential of *K. marxianus* as a robust, resource-efficient cell factory for the valorisation of low nitrogen industrial by-products. In the future, studying the effect of different cost-effective supplements to WP could prove vital for the efficiency of biorefineries based on this feedstock.

References

Ai, Y., Luo, T., Yu, Y., Zhou, J., & Lu, H. (2022). Downregulation of ammonium uptake improves the growth and tolerance of Kluyveromyces marxianus at high temperature. *MicrobiologyOpen*, *11*(3), e1290. https://doi.org/10.1002/MBO3.1290

Belem, M. A. F., & Lee, B. H. (1999). Fed-batch fermentation to produce oligonucleotides from Kluyveromyces marxianus grown on whey. *Process Biochemistry*, *34*(5), 501–509. https://doi.org/10.1016/S0032-9592(98)00118-6

Berg, H. Y., Arju, G., Becerra-Rodríguez, C., Galeote, V., & Nisamedtinov, I. (2023). Unlocking the secrets of peptide transport in wine yeast: insights into oligopeptide transporter functions and nitrogen source preferences. *Applied and Environmental Microbiology*, *89*(11). https://doi.org/10.1128/AEM.01141-23

Brzywczy, J., Natorff, R., Sieńko, M., & Paszewski, A. (2007). Multiple fungal enzymes possess cysteine synthase activity in vitro. *Research in Microbiology*, *158*(5), 428–436. https://doi.org/10.1016/J.RESMIC.2007.03.002

Coral-Medina, A., Fenton, D. A., Varela, J., Camarasa, C., Baranov, P. V., & Morrissey, J. P. (2022). Regulated use of alternative Transcription Start Sites controls the production of cytosolic or mitochondrial forms of branched-chain aminotransferase in Kluyveromyces marxianus. *BioRxiv*, 2022.04.27.489738. https://doi.org/10.1101/2022.04.27.489738

de Lima, E. A., Mandelli, F., Kolling, D., Matsusato Souza, J., de Oliveira Filho, C. A., Ribeiro da Silva, M., Lobo de Mesquita Sampaio, I., Lopes Junqueira, T., Ferreira Chagas, M., Teodoro, J. C., de Morais, E. R., & Murakami, M. T. (2022). Development of an economically competitive Trichoderma-based platform for enzyme production: Bioprocess optimization, pilot plant scale-up, techno-economic analysis and life cycle assessment. *Bioresource*

Technology, *364*, 128019. https://doi.org/10.1016/J.BIORTECH.2022.128019

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15. https://doi.org/10.1093/BIOINFORMATICS/BTS635

Donzella, L., Sousa, M. J., & Morrissey, J. P. (2023). Evolution and functional diversification of yeast sugar transporters. *Essays in Biochemistry*, *67*(5), 811–827. https://doi.org/10.1042/EBC20220233

Doughty, T. W., Domenzain, I., Millan-Oropeza, A., Montini, N., de Groot, P. A., Pereira, R., Nielsen, J., Henry, C., Daran, J. M. G., Siewers, V., & Morrissey, J. P. (2020). Stress-induced expression is enriched for evolutionarily young genes in diverse budding yeasts. *Nature Communications*, *11*(1), 1–9. https://doi.org/10.1038/s41467-020-16073-3

Dubey, P., Ahmad, M. S., Laxman, S., & Bachhawat, A. K. (2025). Seo1p, a high affinity, plasma membrane transporter of the γ-Glu-met dipeptide in yeasts and fungi. *BioRxiv*, 2025.01.15.633159. https://doi.org/10.1101/2025.01.15.633159

Fairbairn, S., McKinnon, A., Musarurwa, H. T., Ferreira, A. C., & Bauer, F. F. (2017). The impact of single amino acids on growth and volatile aroma production by Saccharomyces cerevisiae strains. *Frontiers in Microbiology*, *8*(DEC), 310569. https://doi.org/10.3389/FMICB.2017.02554

Fenton, D. A., Kiniry, S. J., Yordanova, M. M., Baranov, P. V., & Morrissey, J. P. (2022). Development of a ribosome profiling protocol to study translation in Kluyveromyces marxianus. *FEMS Yeast Research*, *22*(1), 1–9. https://doi.org/10.1093/FEMSYR/FOAC024

Fenton, D. A., Świrski, M., O'Connor, P. B. F., Kiniry, S. J., Michel, A. M., Kufel, J., Baranov, P. v, & Morrissey, J. P. (2022). Transcriptional and translational profiling in yeast reveals the use of diverse genome decoding mechanisms

to generate functionally distinct proteoforms. *BioRxiv*, 2022.03.25.485750. https://doi.org/10.1101/2022.03.25.485750

Fonseca, G. G., Heinzle, E., Wittmann, C., Gombert, A. K., Fonseca, G. G., Gombert, A. K., Heinzle, E., & Wittmann, C. (2008). The yeast Kluyveromyces marxianus and its biotechnological potential. *Appl Microbiol Biotechnol*, *79*, 339–354. https://doi.org/10.1007/s00253-008-1458-6

Gabaldón, T. (2020). Hybridization and the origin of new yeast lineages. *FEMS Yeast Research*, *20*(5), 40. https://doi.org/10.1093/FEMSYR/FOAA040

Gao, Y., Gao, L., Kang, Y., Yang, G., Zhao, Z., Zhao, Y., & Li, S. (2024). Non-Targeted Metabolomics Analysis Reveals Metabolite Profiles Change During Whey Fermentation with Kluyveromyces marxianus. *Metabolites*, *14*(12), 694. https://doi.org/10.3390/METABO14120694

Gethins, L., Guneser, O., Demirkol, A., Rea, M. C., Stanton, C., Ross, R. P., Yuceer, Y., & Morrissey, J. P. (2015). Influence of carbon and nitrogen source on production of volatile fragrance and flavour metabolites by the yeast Kluyveromyces marxianus. *Yeast*, *32*(1), 67–76. https://doi.org/10.1002/YEA.3047

Gómez-Márquez, C., Sandoval-Nuñez, D., Gschaedler, A., Romero-Gutiérrez, T., Amaya-Delgado, L., & Alejandro Morales, J. (2022). Diploid genome assembly of Kluyveromyces marxianus NRRL Y-50883 (SLP1). *G3 Genes|Genomes|Genetics*, *12*(1). https://doi.org/10.1093/G3JOURNAL/JKAB347

Groenewald, M., Hittinger, C. T., Bensch, K., Opulente, D. A., Shen, X. X., Li, Y., Liu, C., LaBella, A. L., Zhou, X., Limtong, S., Jindamorakot, S., Gonçalves, P., Robert, V., Wolfe, K. H., Rosa, C. A., Boekhout, T., Čadež, N., Péter, G., Sampaio, J. P., … Rokas, A. (2023). A genome-informed higher rank classification of the biotechnologically important fungal subphylum Saccharomycotina. *Studies in Mycology*, *105*, 1–22. https://doi.org/10.3114/SIM.2023.105.01

Hagman, A., Säll, T., & Piškur, J. (2014). Analysis of the yeast short-term Crabtree effect and its origin. *FEBS Journal*, *281*(21), 4805–4814. https://doi.org/10.1111/febs.13019

Hernández-Mendoza, A. G., Ruiz, H. A., Ortiz-Ceballos, Á. I., Castro-Luna, A. A., Láinez, M., & Martínez-Hernández, S. (2024). Ethanol production from Agave salmiana leaf juices by consolidated bioprocessing comparing two strains of Kluyveromyces marxianus. *Industrial Crops and Products*, *208*, 117839. https://doi.org/10.1016/J.INDCROP.2023.117839

Hittinger, C. T., Steele, J. L., & Ryder, D. S. (2018). Diverse yeasts for diverse fermented beverages and foods. *Current Opinion in Biotechnology*, *49*, 199–206. https://doi.org/10.1016/J.COPBIO.2017.10.004

Hu, N., Yuan, B., Sun, J., Wang, S. A., & Li, F. L. (2012). Thermotolerant Kluyveromyces marxianus and Saccharomyces cerevisiae strains representing potentials for bioethanol production from Jerusalem artichoke by consolidated bioprocessing. *Applied Microbiology and Biotechnology 2012 95:5*, *95*(5), 1359–1368. https://doi.org/10.1007/S00253-012-4240-8

Huang, P., Li, J., Wang, Z., Wang, B., Yuan, L., Zheng, X., & Yang, Z. (2024). Transcriptomic analysis to elucidate the molecular mechanisms underlying the growth advantage of Kluyveromyces marxianus XZ1 and the development of resistance in apple induced by XZ1. *Postharvest Biology and Technology*, *210*, 112775. https://doi.org/10.1016/J.POSTHARVBIO.2024.112775

Imura, M., Nitta, K., Iwakiri, R., Matsuda, F., Shimizu, H., & Fukusaki, E. (2020). Comparison of metabolic profiles of yeasts based on the difference of the Crabtree positive and negative. *Journal of Bioscience and Bioengineering*, *129*(1), 52–58. https://doi.org/10.1016/j.jbiosc.2019.07.007

Inokuma, K., Ishii, J., Hara, K. Y., Mochizuki, M., Hasunuma, T., & Kondo, A. (2016). Complete genome sequence of Kluyveromyces marxianus

NBRC1777, a nonconventional thermotolerant yeast. *Genome Announcements*, *3*(2). https://doi.org/10.1128/GENOMEA.00389-15

Jeong, H., Lee, D. H., Kim, S. H., Kim, H. J., Lee, K., Song, J. Y., Kim, B. K., Sung, B. H., Park, J. C., Sohn, J. H., Koo, H. M., & Kim, J. F. (2012). Genome sequence of the thermotolerant yeast Kluyveromyces marxianus var. marxianus KCTC 17555. *Eukaryotic Cell*, *11*(12), 1584–1585. https://doi.org/10.1128/EC.00260-12

Kassem, J. M. (2015). Future Challenges of Whey Proteins. *International Journal of Dairy Science*, *10*(4), 139–159. https://doi.org/10.3923/ijds.2015.139.159

Kellis, M., Birren, B. W., & Lander, E. S. (2004). Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature*, *428*(6983), 617–624. https://doi.org/10.1038/NATURE02424

Kemsawasd, V., Viana, T., Ardö, Y., & Arneborg, N. (2015). Influence of nitrogen sources on growth and fermentation performance of different wine yeast species during alcoholic fermentation. *Applied Microbiology and Biotechnology*, *99*(23), 10191–10207. https://doi.org/10.1007/S00253-015-6835-3

Kosaka, T., Tsuzuno, T., Nishida, S., Pattanakittivorakul, S., Murata, M., Miyakawa, I., Lertwattanasakul, N., Limtong, S., & Yamada, M. (2022). Distinct Metabolic Flow in Response to Temperature in Thermotolerant Kluyveromyces marxianus. *Applied and Environmental Microbiology*, *88*(6). https://doi.org/10.1128/AEM.02006-21

Koukoumaki, D. I., Papanikolaou, S., Ioannou, Z., Mourtzinos, I., & Sarris, D. (2024). Single-Cell Protein and Ethanol Production of a Newly Isolated Kluyveromyces marxianus Strain through Cheese Whey Valorization. *Foods*, *13*(12), 1892. https://doi.org/10.3390/FOODS13121892

Lane, M. M., Burke, N., Karreman, R., Wolfe, K. H., O'Byrne, C. P., & Morrissey, J. P. (2011). Physiological and metabolic diversity in the yeast

Kluyveromyces marxianus. *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, *100*(4), 507–519. https://doi.org/10.1007/s10482-011-9606-x

Lane, M. M., & Morrissey, J. P. (2010). Kluyveromyces marxianus: A yeast emerging from its sister's shadow. In *Fungal Biology Reviews* (Vol. 24, Issues 1–2, pp. 17–26). Elsevier. https://doi.org/10.1016/j.fbr.2010.01.001

Lee, M. E., DeLoache, W. C., Cervantes, B., & Dueber, J. E. (2015). A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology*, *4*(9), 975–986. https://doi.org/10.1021/sb500366v

Lertwattanasakul, N., Kosaka, T., Hosoyama, A., Suzuki, Y., Rodrussamee, N., Matsutani, M., Murata, M., Fujimoto, N., Suprayogi, Tsuchikane, K., Limtong, S., Fujita, N., & Yamada, M. (2015). Genetic basis of the highly efficient yeast Kluyveromyces marxianus: Complete genome sequence and transcriptome analyses. *Biotechnology for Biofuels*, *8*(1), 1–14. https://doi.org/10.1186/s13068-015-0227-x

Les, E. K., Pometun, E. V., Savin, S. S., Tishkov, V. I., & Pometun, A. A. (2025). Cysteine Synthase: A Key Enzyme of Cysteine Biosynthetic Pathway. *Biochemistry (Moscow) 2025 90:10*, *90*(10), 1311–1331. https://doi.org/10.1134/S0006297925602084

Li, Y., Hou, S., Ren, Z., Fu, S., Wang, S., Chen, M., Dang, Y., Li, H., Li, S., & Li, P. (2023). Transcriptomic analysis reveals hub genes and pathways in response to acetic acid stress in Kluyveromyces marxianus during high-temperature ethanol fermentation. *Stress Biology 2023 3:1*, *3*(1), 1–12. https://doi.org/10.1007/S44154-023-00108-Y

Linder, T. (2019a). A genomic survey of nitrogen assimilation pathways in budding yeasts (sub-phylum Saccharomycotina). *Yeast*, *36*(5), 259–273. https://doi.org/10.1002/YEA.3364

Linder, T. (2019b). Nitrogen assimilation pathways in budding yeasts. In *Non-conventional Yeasts: from Basic Research to Application* (pp. 197–236).

Springer International Publishing. https://doi.org/10.1007/978-3-030-21110-3_7

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 1–21. https://doi.org/10.1186/S13059-014-0550-8

Malina, C., Yu, R., Bjorkeroth, J., Kerkhoven, E. J., & Nielsen, J. (2021). Adaptations in metabolism and protein translation give rise to the Crabtree effect in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(51), e2112836118. https://doi.org/10.1073/PNAS.2112836118

Marcišauskas, S., Ji, B., & Nielsen, J. (2019). Reconstruction and analysis of a Kluyveromyces marxianus genome-scale metabolic model. *BMC Bioinformatics 2019 20:1*, *20*(1), 1–9. https://doi.org/10.1186/S12859-019-3134-5

Martínez-Force, E., & Benitez, T. (1995). Effects of Varying Media, Temperature, and Growth Rates on the Intracellular Concentrations of Yeast Amino Acids. *Biotechnol. Prog*, *11*, 386–392. https://doi.org/10.1021/bp00034a004

Matsuda, F., Kamiyama, A., Yamasaki, K., Seike, T., & Okahashi, N. (2025). ΔΔG Method for Elucidating Key Control Reactions from Relative Quantification Metabolome Data: Comparative Analysis of Yeast Glycolysis. *Analytical Chemistry*, *97*(12), 6391–6398. https://doi.org/10.1021/ACS.ANALCHEM.4C04480

Mejía-Barajas, J., Montoya-Pérez, R., Manzo-Avalos, S., Cortés-Rojo, C., Riveros-Rosas, H., Cervantes, C., & Saavedra-Molina, A. (2018). Fatty acid addition and thermotolerance of Kluyveromyces marxianus. *FEMS Microbiology Letters*, *365*(7), 43. https://doi.org/10.1093/FEMSLE/FNY043

Merico, A., Sulo, P., Piškur, J., & Compagno, C. (2007). Fermentative lifestyle in yeasts belonging to the Saccharomyces complex. *FEBS Journal*, *274*(4), 976–989. https://doi.org/10.1111/j.1742-4658.2007.05645.x

Montini, N., Doughty, T. W., Domenzain, I., Fenton, D. A., Baranov, P. V., Harrington, R., Nielsen, J., Siewers, V., & Morrissey, J. P. (2022). Identification of a novel gene required for competitive growth at high temperature in the thermotolerant yeast Kluyveromyces marxianus. *Microbiology (Reading, England)*, *168*(3). https://doi.org/10.1099/MIC.0.001148

Moreno, F., & Herrero, P. (2002). The hexokinase 2-dependent glucose signal transduction pathway of Saccharomyces cerevisiae. *FEMS Microbiology Reviews*, *26*(1), 83–90. https://doi.org/10.1111/J.1574-6976.2002.TB00600.X

Nurcholis, M., Nitiyon, S., Suprayogi, Rodrussamee, N., Lertwattanasakul, N., Limtong, S., Kosaka, T., & Yamada, M. (2019). Functional analysis of Mig1 and Rag5 as expressional regulators in thermotolerant yeast Kluyveromyces marxianus. *Applied Microbiology and Biotechnology*, *103*(1), 395–410. https://doi.org/10.1007/S00253-018-9462-Y

O'Donoghue, L. T., & Murphy, E. G. (2023). Nondairy food applications of whey and milk permeates: Direct and indirect uses. *Comprehensive Reviews in Food Science and Food Safety*, *22*(4), 2652–2677. https://doi.org/10.1111/1541-4337.13157

Ortiz-Merino, R. A., Varela, J. A., Coughlan, A. Y., Hoshida, H., da Silveira, W. B., Wilde, C., Kuijpers, N. G. A., Geertman, J. M., Wolfe, K. H., & Morrissey, J. P. (2018). Ploidy variation in Kluyveromyces marxianus separates dairy and non-dairy isolates. *Frontiers in Genetics*, *9*(MAR), 94. https://doi.org/10.3389/fgene.2018.00094

Parekh, S., & Margaritis, A. (1986). Production of inulinase (β-fructan fructanohydrolase) by kluyveromyces marxianus. *Agricultural and Biological Chemistry*, *50*(4), 1085–1087. https://doi.org/10.1080/00021369.1986.10867525

Parente, D. C., Cajueiro, D. B. B., Moreno, I. C. P., Leite, F. C. B., De Barros Pita, W., & De Morais, M. A. (2018). On the catabolism of amino acids in the

yeast Dekkera bruxellensis and the implications for industrial fermentation processes. *Yeast*, *35*(3), 299–309. https://doi.org/10.1002/YEA.3290

Peng, Y., Gao, C., Wang, S., Ozaki, M., & Takigawa, A. (2003). Non-filamentous sludge bulking caused by a deficiency of nitrogen in industrial wastewater treatment. *Water Science and Technology*, *47*(11), 289–295. https://doi.org/10.2166/WST.2003.0617

Pinheiro, R., Belo, I., & Mota, M. (2003). Growth and β-galactosidase activity in cultures of Kluyveromyces marxianus under increased air pressure. *Letters in Applied Microbiology*, *37*(6), 438–442. https://doi.org/10.1046/j.1472-765X.2003.01429.x

Pollegioni, L., Diederichs, K., Molla, G., Umhau, S., Welte, W., Ghisla, S., & Pilone, M. S. (2002). Yeast d-Amino Acid Oxidase: Structural Basis of its Catalytic Properties. *Journal of Molecular Biology*, *324*(3), 535–546. https://doi.org/10.1016/S0022-2836(02)01062-8

Rajkumar, A. S., & Morrissey, J. P. (2022). Protocols for marker-free gene knock-out and knock-down in Kluyveromyces marxianus using CRISPR/Cas9. *FEMS Yeast Research*, *22*(1), 67. https://doi.org/10.1093/FEMSYR/FOAB067

Rajkumar, A. S., Varela, J. A., Juergens, H., Daran, J. M. G., & Morrissey, J. P. (2019). Biological parts for Kluyveromyces marxianus synthetic biology. *Frontiers in Bioengineering and Biotechnology*, *7*(MAY). https://doi.org/10.3389/fbioe.2019.00097

Raposo, S., Constantino, A., Rodrigues, F., Rodrigues, B., & Lima-Costa, M. E. (2016). Nitrogen Sources Screening for Ethanol Production Using Carob Industrial Wastes. *Applied Biochemistry and Biotechnology 2016 181:2*, *181*(2), 827–843. https://doi.org/10.1007/S12010-016-2252-Z

Roca-mesa, H., Sendra, S., Mas, A., Beltran, G., & Torija, M. J. (2020). Nitrogen Preferences during Alcoholic Fermentation of Different Non-Saccharomyces Yeasts of Oenological Interest. *Microorganisms 2020, Vol.*

8, Page 157, *8*(2), 157. https://doi.org/10.3390/MICROORGANISMS8020157

Rojo, M. C., Talia, P. M., Lerena, M. C., Ponsone, M. L., Gonzalez, M. L., Becerra, L. M., Mercado, L. A., Martín-Arranz, V., Rodríguez-Gómez, F., Arroyo-López, F. N., & Combina, M. (2023). Evaluation of different nitrogen sources on growth and fermentation performance for enhancing ethanol production by wine yeasts. *Heliyon*, *9*(12), e22608. https://doi.org/10.1016/j.heliyon.2023.e22608

Rollero, S., Bloem, A., Ortiz-Julien, A., Bauer, F. F., Camarasa, C., & Divol, B. (2019). A comparison of the nitrogen metabolic networks of Kluyveromyces marxianus and Saccharomyces cerevisiae. *Environmental Microbiology*, *21*(11), 4076–4091. https://doi.org/10.1111/1462-2920.14756

Romagnoli, G., Verhoeven, M. D., Mans, R., Fleury Rey, Y., Bel-Rhlid, R., van den Broek, M., Maleki Seifar, R., Ten Pierick, A., Thompson, M., Müller, V., Wahl, S. A., Pronk, J. T., & Daran, J. M. (2014). An alternative, arginase-independent pathway for arginine metabolism in Kluyveromyces lactis involves guanidinobutyrase as a key enzyme. *Molecular Microbiology*, *93*(2), 369–389. https://doi.org/10.1111/MMI.12666

Roy, P., Gahlawat, V. K., Saravanan, C., & Singh, B. P. (2023). Enhancing bioflavor production by solid-state fermentation using Kluyveromyces marxianus and l-phenylalanine. *Journal of Basic Microbiology*, *63*(1), 75–91. https://doi.org/10.1002/JOBM.202200503

Sakihama, Y., Hidese, R., Hasunuma, T., & Kondo, A. (2019). Increased flux in acetyl-CoA synthetic pathway and TCA cycle of Kluyveromyces marxianus under respiratory conditions. *Scientific Reports 2019 9:1*, *9*(1), 5319-. https://doi.org/10.1038/s41598-019-41863-1

Schabort, D. T. W. P., Letebele, P., Steyn, L., Kilian, S. G., & Du Preez, J. C. (2016). Differential RNA-seq, Multi-Network Analysis and Metabolic Regulation Analysis of Kluyveromyces marxianus Reveals a Compartmentalised

Response to Xylose. *PLOS ONE*, *11*(6), e0156242. https://doi.org/10.1371/JOURNAL.PONE.0156242

Shen, X. X., Steenwyk, J. L., LaBella, A. L., Opulente, D. A., Zhou, X., Kominek, J., Li, Y., Groenewald, M., Hittinger, C. T., & Rokas, A. (2020). Genome-scale phylogeny and contrasting modes of genome evolution in the fungal phylum Ascomycota. *Science Advances*, *6*(45). https://doi.org/10.1126/SCIADV.ABD0079

Smets, J., Godoy, H. E., Goossenaerts, J., Van Bun, E., Deparis, Q., Bauwens, J., Ortiz-Merino, R. A., Mancera, E., DeLuna, A., & Verstrepen, K. J. (2025). Metabolic engineering and adaptive laboratory evolution of Kluyveromyces Marxianus for lactic acid production. *Microbial Cell Factories*, *24*(1), 1–17. https://doi.org/10.1186/S12934-025-02805-X

Steenwyk, J. L., & Rokas, A. (2018). Copy number variation in fungi and its implications for wine yeast genetic diversity and adaptation. *Frontiers in Microbiology*, *9*(FEB), 340048. https://doi.org/10.3389/FMICB.2018.00288

Tsermoula, P., Bechshøft, M. R., Friis, C., Engelsen, S. B., & Khakimov, B. (2023). Screening of non-protein nitrogen compounds in lactose refining streams from industrial whey permeate processing. *Food Chemistry*, *405*, 134716. https://doi.org/10.1016/J.FOODCHEM.2022.134716

Varela, J. A., Puricelli, M., Montini, N., & Morrissey, J. P. (2019). Expansion and diversification of MFS transporters in Kluyveromyces marxianus. *Frontiers in Microbiology*, *10*(JAN). https://doi.org/10.3389/fmicb.2018.03330

Varela, J. A., Puricelli, M., Ortiz-Merino, R. A., Giacomobono, R., Braun-Galleani, S., Wolfe, K. H., & Morrissey, J. P. (2019). Origin of Lactose Fermentation in Kluyveromyces lactis by Interspecies Transfer of a Neo-functionalized Gene Cluster during Domestication. *Current Biology*, *29*(24), 4284-4290.e2. https://doi.org/10.1016/j.cub.2019.10.044

Walker, G. M. (2014). *Media for Industrial Fermentations*. https://doi.org/10.1016/B978-0-12-384730-0.00107-5

Walvekar, A. S., & Laxman, S. (2019). Methionine at the Heart of Anabolism and Signaling: Perspectives From Budding Yeast. *Frontiers in Microbiology*, *10*, 499926. https://doi.org/10.3389/FMICB.2019.02624

Walvekar, A. S., Srinivasan, R., Gupta, R., & Laxman, S. (2018). Methionine coordinates a hierarchically organized anabolic program enabling proliferation. *Molecular Biology of the Cell*, *29*(26), 3183–3200. https://doi.org/10.1091/MBC.E18-08-0515

Wolfe, K. H. (2015). Origin of the Yeast Whole-Genome Duplication. *PLOS Biology*, *13*(8), e1002221. https://doi.org/10.1371/JOURNAL.PBIO.1002221

Yu, Y., Mo, W., Ren, H., Yang, X., Lu, W., Luo, T., Zeng, J., Zhou, J., Qi, J., & Lu, H. (2021). Comparative Genomic and Transcriptomic Analysis Reveals Specific Features of Gene Regulation in Kluyveromyces marxianus. *Frontiers in Microbiology*, *12*, 404. https://doi.org/10.3389/fmicb.2021.598060

Zhang, B., Zhu, Y., Zhang, J., Wang, D., Sun, L., & Hong, J. (2017). Engineered Kluyveromyces marxianus for pyruvate production at elevated temperature with simultaneous consumption of xylose and glucose. *Bioresource Technology*, *224*, 553–562. https://doi.org/10.1016/j.biortech.2016.11.110

Zhang, J., Xu, T., Wang, X., Jing, X., Zhang, J., Hong, J., Xu, J., & Wang, J. (2022). Lignocellulosic xylitol production from corncob using engineered Kluyveromyces marxianus. *Frontiers in Bioengineering and Biotechnology*, *10*, 1029203. https://doi.org/10.3389/FBIOE.2022.1029203

Zhang, Z., Lan, Q., Yu, Y., Zhou, J., & Lu, H. (2022). Comparative metabolome and transcriptome analyses of the properties of Kluyveromyces marxianus and Saccharomyces yeasts in apple cider fermentation. *Food Chemistry: Molecular Sciences*, *4*, 100095. https://doi.org/10.1016/J.FOCHMS.2022.100095

Supplementary files

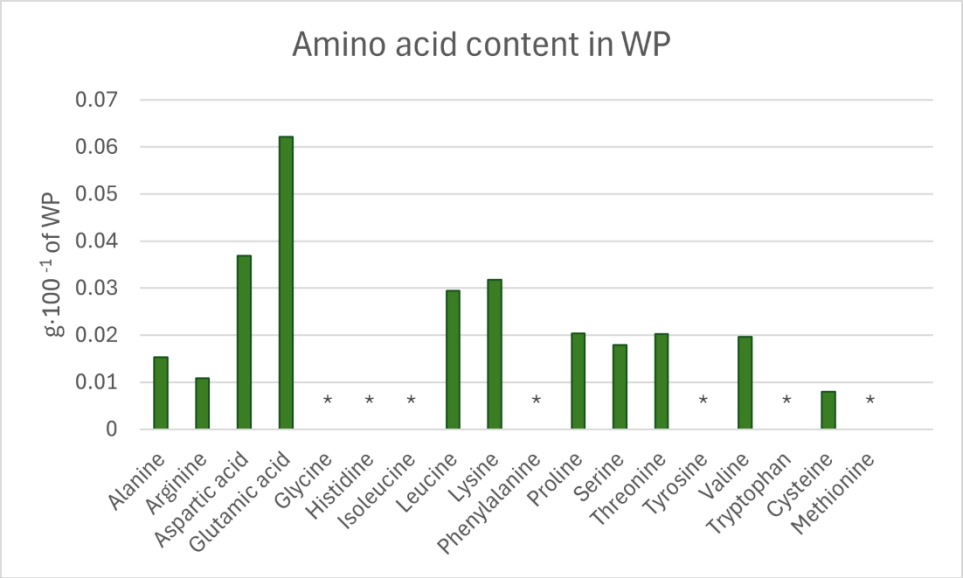


Fig. S1. Amino acid content in WP. Content of each single amino acids in WP. * Indicates below limit of detection. Limit detection for Glycine and Histidine < 0.02 g·100g^{-1} WP. Isoleucine < 0.035 g·100g^{-1} WP. Phenylalanine < 0.031 g·100g^{-1} WP. Tyrosine < 0.023 g·100g^{-1} WP. Tryptophan < 0.01 g·100g^{-1} WP. Methionine < 0.024 g·100g^{-1} WP.

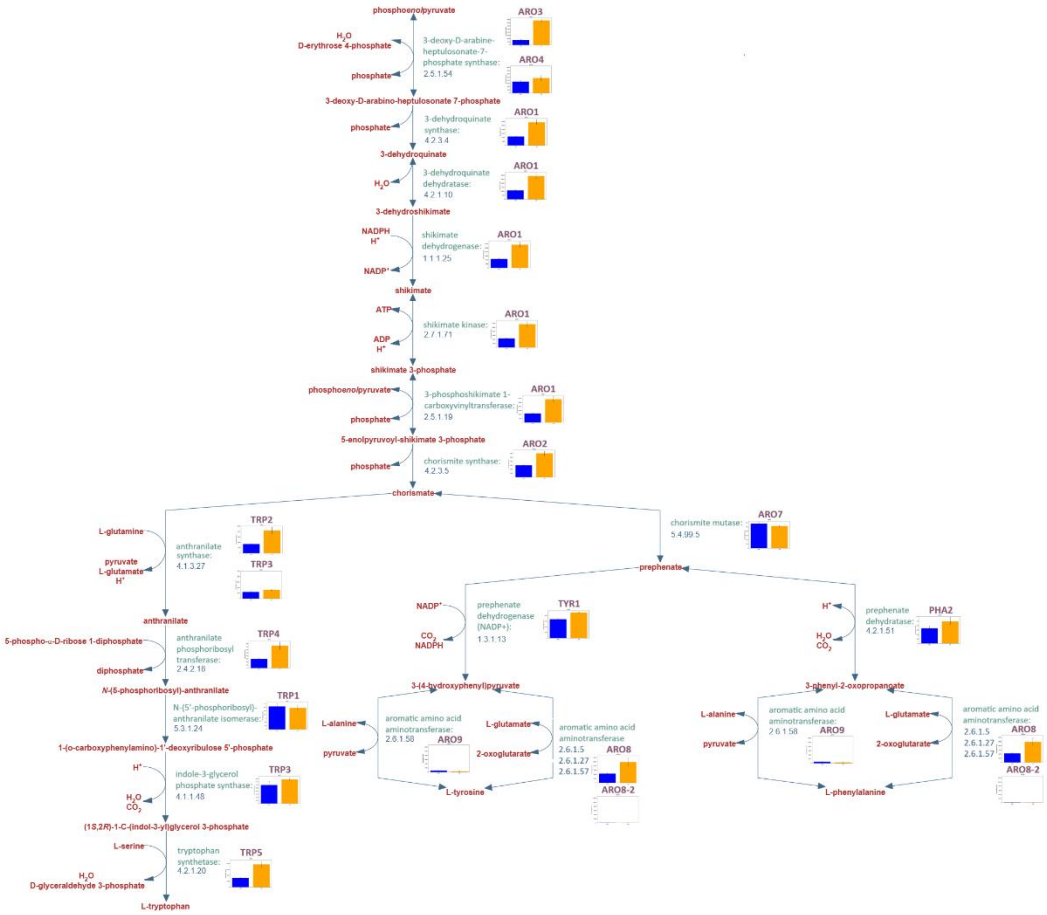


Fig. S2. Aromatic amino acid synthesis in *K. marxianus*. Biosynthetic pathway in *K. marxianus* for production of aromatic amino acids from Pentose Phosphate Pathway precursors. Each enzymatic step is annotated with a corresponding histogram representing the expression level of every enzyme under two different conditions, WPT in blue, WP in orange. When more than one enzyme is known to catalyse the same or the reverse step, their level of expression can be compared visually. The histograms provide a comparative visualisation of transcriptional regulation and metabolic fluxes across the pathway.

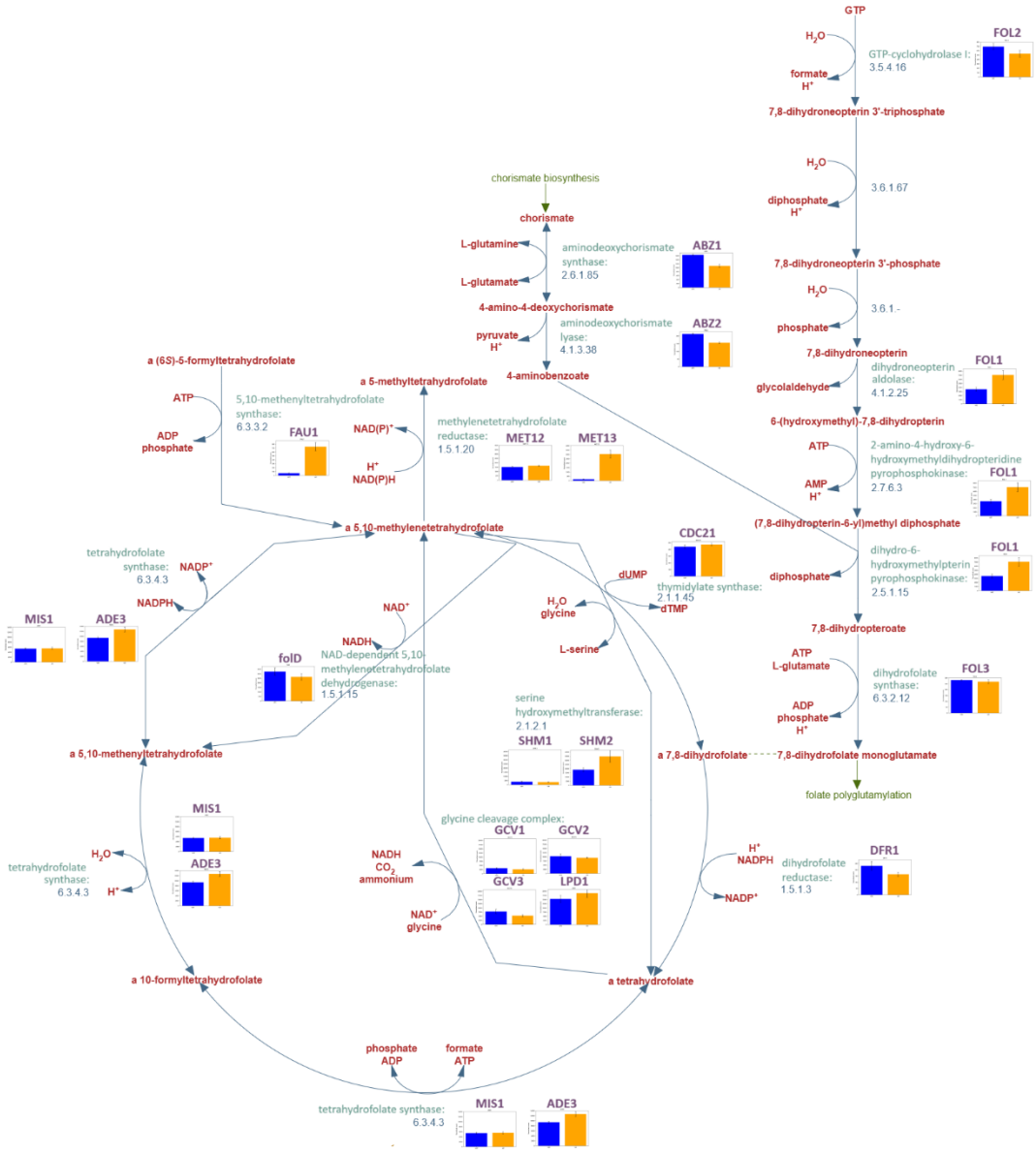


Fig. S3. Tetrahydrofolate synthesis in *K. marxianus*. Tetrahydrofolate synthesis pathway in *K. marxianus* from its different precursors chorismate and guanine triphosphate (GPT). Each enzymatic step is annotated with a corresponding histogram representing the expression level of every enzyme under two different conditions, WPT in blue, Whey Permeate WP in orange. When more than one enzyme is known to catalyse the same or the reverse step, their level of expression can be compared visually. The histograms provide a comparative visualisation of transcriptional regulation and metabolic fluxes across the pathway.

Table S1. List of all the expressed genes in *K. marxianus* during growth on WP versus WPT. Positive $\log_2$ fold change values indicate higher expression in WP, while negative values indicate higher expression in WPT. Biological process was assigned based on annotation from UniProt and was manually curated using the GWIPS-viz browser, synteny tools, AlphaFold and SeekFold. This data is available on the Zenodo link provided in the materials and methods section.

Table S2. List of number of transcripts in each of the three replicates and average of the three during growth on WP and WPT. Genes are grouped by metabolically relevant pathway. Biological process was assigned based on annotation from UniProt and was manually curated using the GWIPS-viz browser, synteny tools, AlphaFold and SeekFold. Table S2 is available under request and will be soon available on Zenodo.

K. marxianus **DMKU3-1042**	***K. marxianus*** **CBS 6556**	***K. lactis*** **CLIB 210**	***S. cerevisiae*** **S288C**	**Possible function**
KLMX_50200	KMXK_0E02080	KLLA0B05995g	X	Cell wall organisation. Similarity with HRQ2 from *Candida albicans*.
KLMX_30014	KMXK_0C00160	X	X	Transcription factor containing Zinc Cluster Factor 32 domain from Zn(2)-C6 family.
KLMX_40008	KMXK_0D00220	X	X	Transcription factor belonging to Zn(2)-C6 family. Probable N-regulator (similarities with THI1, PUT3 and ASG1).
KLMX_70318	KMXK_0H03250	KLLA0B11880g	MET28	Transcription factor activates methionine biosynthesis.
KLMX_60300	KMXK_0F02750	KLLA0F25190g	X	Transcription activator, very similar to GCR2. Paralog that controls other metabolic pathway.
KLMX_30322	KMXK_0C03260	KLLA0B10373g	SPO23	Confirmed by synteny analysis to be SPO23.

KLMX_80401	KMXK_0G04100	KLLA0E21319g	YHR112C	Trans-sulphurase. Possible cystathionine beta-lyase. YHR112C *K. marxianus* ortholog.
KLMX_20060	KMXK_0B00690	KLLA0E10935g	X	URO1 urate oxidase.
KLMX_50443	KMXK_0E04500	KLLA0D15356g	X	Transcription activator. Paralog of DAL81.
KLMX_40310	KMXK_0D03250	KLLA0F26686g	YKL068W-A	*K. marxianus* YKL068W-A ortholog. Unknown, possible role in signal transduction.
KLMX_10167	KMXK_0A06850	KLLA0C03784g	SPO21	Confirmed by synteny analysis to be SPO21.
KLMX_40637	KMXK_0D06550	KLLA0B00495g	YGL117W	Recently proposed to be ARO5: negatively regulates the feedback inhibition of ARO3 by phenylalanine.
KLMX_10836	KMXK_0A00170	X	X	Possible Cell wall protein with similarities to other cell wall proteins (CWP1, PIR1, BUT2).
KLMX_70437	X	X	X	FLO9-like.

Table S3. List of unannotated genes that show significant differential expression during growth on WP versus WPT. Genome annotation in DMKU3-1042 was included and ortholog in other relevant yeasts, *K. marxianus* (CBS 6556), *K. lactis* (CLIB 210) and *S. cerevisiae* (S288C), were added when possible. Function was assigned based on annotation from UniProt and was manually curated using the GWIPS-viz browser, synteny tools, AlphaFold and SeekFold.

Script S1. Python script used for the transcriptome analysis. The script processes RNA-seq gene count files generated by STAR to construct a gene expression matrix. Reverse-strand read counts are extracted from multiple samples and combined into a single unnormalised count matrix. Genes corresponding to non-coding or ribosomal features are filtered out. The resulting matrix is exported for downstream analysis. Additionally, a normalised count matrix is imported and used to compute pairwise Spearman correlations between samples, which are visualised as a heatmap to assess sample similarity.

```
import pandas as pd

import os

import seaborn as sns

import matplotlib.pyplot as plt

import numpy as np


plt.style.use('default')


gene_keys = {}


with open('gene_keys.tsv', 'r') as infile:
    for line in infile:
        line = line.strip().replace('KLMA_', 'KLMX_')
        gene_id, gene_name = line.split('\t')
        gene_keys[gene_id] = gene_name


# ---------------- Build raw count matrix ----------------


df = pd.read_csv(
    'PATH_TO_FIRST_FILE',
    sep='\t',
    names=['gene_id', 'unstranded', '+', '-']
)


df = df.iloc[4:]  # STAR metadata rows

df = df.drop(columns=['unstranded', '+', '-'])


# ---------------- Add sample columns ----------------
```

```
for file in os.listdir('FOLDER_PATH'):

    # Keep your original logic (dataset-specific)
    if file.endswith('C2'):

        sample_id = file.split('ReadsPer')[0]

        if '-' in sample_id:
            sample_id = sample_id.split('_')[0]

        print(sample_id)

        dfx = pd.read_csv(
            os.path.join('FOLDER_PATH', file),
            sep='\t',
            names=['gene_id', 'unstranded', '+', '-']
        )

        dfx = dfx.iloc[4:]  # STAR metadata rows

        df[sample_id] = dfx['-']

# ---------------- Final formatting ----------------
df.index = df['gene_id']
df = df.drop(columns=['gene_id'])

# Experimental sample order
df = df[
    ['CR_1_S1', 'CR_3_S2', 'CR_4_S3',
     'CR_AWTE1_S4', 'CR_AWTE3_S5', 'CR_AWTE4_S6',
     'CR_CWE1_S7', 'CR_CWE3_S8', 'CR_CWE4_S9']
]
```

```
# ---------------- Remove unwanted gene types ----------------

df_unnormalized = df.copy()


# Vectorised filtering (safer than iterrows)

df_unnormalized = df_unnormalized[

    ~df_unnormalized.index.str.contains('KLMX_R|ENSRNA')

]


# ---------------- Save output ----------------

df_unnormalized.to_csv('unnormalized.tsv', sep='\t')


df_norm = pd.read_csv('normalized.tsv', sep='\t', index_col=0)


# ---------------- Correlation ----------------

r = df_norm.corr(method="spearman")


plt.figure(dpi=300, figsize=(6, 5))

sns.heatmap(r,    linewidth=0.5,    cbar_kws={'label':    'Spearman
correlation'})

plt.title('RNA-seq normalized count correlation')

plt.show()
```

Script S2. Python script used for the visualisation of the transcriptome analysis. The script generates a volcano plot from differential gene expression data and highlights significantly regulated genes belonging to predefined metabolic pathways. Genes are filtered based on a significance threshold (adjusted p-value), and selected functional groups are colour-coded to facilitate biological interpretation.

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
plt.rcParams.update({
    'font.size': 10,  # Change this value to adjust font size
    'axes.labelsize': 20,  # Font size of x and y labels
    'axes.titlesize': 20,  # Font size of subplot titles
    'xtick.labelsize': 20,  # Font size of x-axis tick labels
    'ytick.labelsize': 20,  # Font size of y-axis tick labels
    'legend.fontsize': 20,  # Font size of legend
    'font.family' : "Tw Cen MT"
})
# Read the Excel file and extract the required columns
data = pd.read_excel('')
gene_names = data.iloc[:, 1]
fold_changes = data.iloc[:, 3]
p_values = data.iloc[:, 7]

# Set a threshold for significance
p_value_threshold = 10**-25

plt.scatter(-fold_changes, -1 * np.log10(p_values), color='gray')
L1_fc = []
L1_pval = []
L2_fc = []
L2_pval = []
L3_fc = []
L3_pval = []
```

```
L4_fc = []

L4_pval = []

L5_fc = []

L5_pval = []

L6_fc = []

L6_pval = []

L7_fc = []

L7_pval = []

L8_fc = []

L8_pval = []


Lys_biosyn = ["LYS1", "LYS2", "LYS4", "LYS9", "LYS12", "LYS21"]

Arg_biosyn = ["ARG2", "ARG4", "ARG7", "ARG8", "CAR2"]

BCA_biosyn = ["ILV2", "ILV3", "ILV5", "ILV6", "LEU1", "LEU2",
"LEU4"]

Glyox = ["AGX1", "IDP1", "LEU2"]

S_uptake = ["MET1", "MET2", "MET3", "MET5","MET10", "MET14",
"MET16", "MET22", "KLMX_70318", "MET32", "SUL2"]

Met_biosyn = ["HOM2", "HOM3", "str3", "MET6", "MET13", "MET17",
"MUP1", "KLMX_80401", "cysB"]

N_uptake = ["DUR3", "SEO1", "GAP1", "MEP2"]

Cu_uptake = ["ARO10", "DUR1,2", "PDC1", "FRE1", "GDH3", "GLT1",
"MEP2", "MET6", "SUN4"]


for fc, p, gene in zip(fold_changes, p_values, gene_names):

    if p < p_value_threshold and gene in Lys_biosyn:

        L1_fc.append(-fc)

        L1_pval.append(-1 * np.log10(p))

    if p < p_value_threshold and gene in Arg_biosyn:

        L2_fc.append(-fc)

        L2_pval.append(-1 * np.log10(p))

    if p < p_value_threshold and gene in BCA_biosyn:

        L3_fc.append(-fc)

        L3_pval.append(-1 * np.log10(p))
```

```
    if p < p_value_threshold and gene in Glyox:
        L4_fc.append(-fc)
        L4_pval.append(-1 * np.log10(p))
    if p < p_value_threshold and gene in Met_biosyn:
        L5_fc.append(-fc)
        L5_pval.append(-1 * np.log10(p))
    if p < p_value_threshold and gene in S_uptake:
        L6_fc.append(-fc)
        L6_pval.append(-1 * np.log10(p))
    if p < p_value_threshold and gene in N_uptake:
        L7_fc.append(-fc)
        L7_pval.append(-1 * np.log10(p))
    if p < p_value_threshold and gene in Cu_uptake:
        L8_fc.append(-fc)
        L8_pval.append(-1 * np.log10(p))
    else:
        continue


plt.scatter(-fold_changes, -1 * np.log10(p_values), color='gray')
plt.scatter(L1_fc,    L1_pval,    color='green',    label='Lysine
biosynthesis')
plt.scatter(L2_fc,  L2_pval,  color='lightgreen',  label='Arginine
biosynthesis')
plt.scatter(L3_fc,  L3_pval,  color='cyan',  label='Branched  chain
\namino acid biosynthesis')
plt.scatter(L4_fc, L4_pval, color='violet', label='Glyoxylate and
Alanine')
plt.scatter(L5_fc,   L5_pval,   color='yellow',   label='Methionine
biosynthesis')
plt.scatter(L6_fc, L6_pval, color='red', label='Sulphur uptake')
plt.scatter(L7_fc, L7_pval, color='blue', label='Nitrogen uptake')
plt.scatter(L8_fc,  L8_pval,  color='orange',  label='Copper  ion
uptake')


# Add labels for significant genes
```

```
significant_genes = []

for fc, p, gene in zip(fold_changes, p_values, gene_names):

   if p < p_value_threshold:

      plt.annotate(gene, (-fc, -1 * np.log10(p)), textcoords='offset
points', xytext=(5, -5), ha='center')

       significant_genes.append(gene)


# Add threshold lines

plt.axhline(-1    *    np.log10(p_value_threshold),    color='red',
linestyle='--', linewidth=0.5)

plt.axhline(-1  *  np.log10(0.01),  color='black',  linestyle='--',
linewidth=0.5)


# Set plot title and labels

plt.title('WPTryptone vs. WP')

plt.xlabel('log2 fold change', fontsize=20)

plt.ylabel('- log10 p-value', fontsize=20)

#plt.tick_params(axis='x', labelsize=16)

#plt.tick_params(axis='y', labelsize=16)

# Show the legend

plt.legend(fontsize=20, loc="upper left")


# Display the plot

plt.show()
```

Chapter 3

MoClo-compatible secretion expansion and Yeast Surface Display platform for *Kluyveromyces marxianus*

This chapter is under preparation for submission to the ACS Synthetic Biology

Abstract

K. marxianus is emerging as a superior alternative to traditional yeast hosts due to its rapid growth and thermotolerance; however, the lack of standardised genetic tools for protein secretion hinders its widespread industrial adoption. To bridge this gap, we developed a comprehensive, Yeast Tool Kit-compatible secretion expansion and a Yeast Surface Display platform. By combining *in silico* secretome mapping with a Design of Experiments approach, we identified and characterised a diverse library of native secretion signals. We integrated these parts into a modular cloning workflow and validated their performance using enzymatic assays and our novel platform across three distinct protein classes: an enzyme (Mel1p), a fluorescent reporter (Venus), and a biopharmaceutical precursor (Csf2p). Notably, we identified native signals that significantly outperform the commonly used *S. cerevisiae* α-mating factor. This work provides the synthetic biology community with a robust, characterised toolkit that includes secretion signals, reporters, and surface display components, streamlining the engineering of *K. marxianus* and establishing it as a highly competitive chassis for next-generation recombinant protein production.

Introduction

The global demand for protein is increasing, in particular for the production of biopharmaceuticals, biocatalysts and food-protein alternatives (Balasubramanian et al., 2021; Fasim et al., 2021; Ghag et al., 2021; Hong et al., 2022; Idiris et al., 2010; Jayakrishnan et al., 2024; Li et al., 2022; Liu et al., 2023; Nadar et al., 2024). In the biotechnology sector, these proteins are usually produced via heterologous expression. Nevertheless, many of the proteins of interest are naturally produced intracellularly, and cell lysis or protein solubilisation would be required prior to purification, complicating recovery of the product (Bhatwa et al., 2021; Gao et al., 2013; Salazar & Asenjo, 2007). Protein secretion is desirable for heterologous protein production since it simplifies the purification steps, rending bioprocesses more economically feasible (Cleaver et al., 2025; Dupuis et al., 2023; Madhavan et al., 2021). To produce heterologous proteins, a number of microbial hosts are generally used: the bacteria, *E. coli*, *Bacillus subtilis*; the yeasts, *K. phaffii*, *S. cerevisiae;* and fungi such as *Aspergillus* spp. are common choices in the industry due to the extensive knowledge, available tools and standardised protocols and procedures (Biermann et al., 2025; Bonanni et al., 2023; Claes et al., 2024; Kittler et al., 2025; Köppl et al., 2022; Li et al., 2022; Liu et al., 2023; Ntana et al., 2020; Ziegler et al., 2021). Nevertheless, although well studied, these traditional workhorses for the biotechnology sector do not suit every operation and other production hosts could have more potential for specific applications.

To express heterologous proteins in these microbial platforms, an expression vector is usually assembled: a plasmid genetically engineered to enable the transcription and translation of a specific gene of interest. Researchers working with these platforms have access to different molecular biology tools that allow the construction of expression cassettes and strains (Caro-Astorga et al., 2025; Lee et al., 2015; Sarkari et al., 2017; Tong et al., 2021). These vectors contain essential regulatory elements, such as promoters and terminators, allowing for controlled and efficient expression of recombinant proteins. To achieve assembly of expression vectors, different methods can be applied (Engler et al., 2008; Gibson et al., 2009; Kouprina & Larionov, 2006; Walhout et al., 2000). One of the most popular cloning strategies is Golden Gate (GG),

also known as MoClo. This method, which relies on type IIs restriction enzymes (Szybalski et al., 1991), allows for one step *in vitro* plasmid assembly with high throughput capability. Many research groups have developed molecular biology toolkits that comprise standardised genetic parts such as promoters and terminators that are ready to assemble with the coding sequence of a protein of interest (Caro-Astorga et al., 2025; Crozet et al., 2018; Engler et al., 2014; Fonseca et al., 2019). In particular for yeast, the YTK developed for *S. cerevisiae* is successful due to the level of characterisation, completeness and ease of use (Lee et al., 2015). Since its development many groups have contributed by adding more parts, expanding its complexity level and capabilities for *S. cerevisiae* and other yeasts (Celińska et al., 2017; Otto et al., 2021; Rajkumar et al., 2019; Shaw et al., 2023). However, although YTK allows for protein fusions that could give rise to secreted proteins, it was not designed for the construction of secreted proteins. Some research groups addressed this by developing YTK-compatible expansions that allow secretion in *S. cerevisiae* or other yeasts like *K. phaffii* or *Y. lipolytica* (Celińska et al., 2018; Obst et al., 2017; O'Riordan et al., 2024).

K. marxianus stands out as a versatile yeast for the production of heterologous proteins (Baptista & Domingues, 2022; Bilal et al., 2022; Lane & Morrissey, 2010) because is a fast-growing yeast that is able to consume different carbon sources, is thermotolerant, and prefers aerobic respiration rather than fermentation, which is a desired trait when producing heterologous proteins (Gomes et al., 2018; Hagman et al., 2014; Lane & Morrissey, 2010). It is qualified to produce food proteins due to its Qualified Presumption Safety (QPS) by the European Food Safety Authority (EFSA) and Generally Recognized As Safe (GRAS) status by U.S. Food and Drug Administration (FDA). Moreover, its phylogenetical proximity to *S. cerevisiae* allows the adoption of many tools and knowledge from this widely studied yeast (Knoshaug et al., 2015; Li et al., 2017; Linder, 2019; Rodrigues et al., 2006). All these traits make *K. marxianus* a good alternative for heterologous protein expression and secretion for industrial applications. Furthermore, *K. marxianus* has been quite well characterised and molecular biology tools like CRISPR/Cas9 or MoClo DNA assembly toolkits are available (Domenzain et al., 2022; Donzella et al., 2021; Gombert et al., 2016; Kumar

et al., 2021; Rajkumar et al., 2019). However, secretion of heterologous proteins is not well characterised, few tools are available for this yeast, and these are not fully compatible with other well-established toolkits (An et al., 2024; Bassett et al., 2024).

In *S. cerevisiae*, two main pathways for protein secretion exist: the co-translational pathway, which is a passive process based on Brownian motion to translocate the nascent polypeptide from the cytoplasm to the endoplasmic reticulum (ER) lumen, and the post-translational pathway which requires different chaperones for translocation and folding, resulting in a much more complex process (Delic et al., 2013; Nyathi et al., 2013). Both pathways start by the identification of an N-terminal peptide known as secretion signal (SS) that precedes the protein to be secreted. An SS is typically composed of a positively charged N-terminal amino acid sequence, a hydrophobic core and a polar, uncharged C-terminal domain, which contains a peptidase cleavage site (Celińska et al., 2018; Cleaver et al., 2025; Owji et al., 2018; Yarimizu et al., 2015). In the co-translational pathway, the SS from the nascent polypeptide is identified by the Signal Recognition Particle and carried to the endoplasmic ER membrane, where it is translocated into the ER lumen through the translocon complex. The SS is then cleaved, releasing the secreted protein into the ER lumen for trafficking into the secretion pathway via the Golgi apparatus (Delic et al., 2013; Nyathi et al., 2013; Zahrl et al., 2022). In contrast, the post-translational secretion pathway is commonly directed by SSs that contain a more complex “pre-pro” structure. The “pre” region is analogue to the SSs directing co-translational secretion, while the “pro” region contains internal protease cleavage sites for subsequent processing during secretion. “Pro” region moieties can vary greatly in sequence, but they enable translocation into the ER by interacting with chaperones, which aid to reach proper protein folding (Chaudhuri et al., 1992; Cleaver et al., 2025; Fitzgerald & Glick, 2014). Although the structure of an SS is well-known, the efficiency to secrete a given protein is difficult to predict, and it also depends on features of the protein to be synthetised such as structure or size (Owji et al., 2018; Xue et al., 2023; Yarimizu et al., 2015; Zsebo et al., 1986). Although there is high natural SS diversity, most heterologous protein secretion systems rely on a few well-characterised SSs (Aza et al., 2021; Gomes et al., 2018; Obst et al., 2017; Xue et al., 2023).

Evaluation of the effectiveness of SSs typically requires measuring level of secreted protein. This can be direct visualisation or by measuring the activity of the protein however, non-enzymatic proteins are often challenging to detect, requiring indirect strategies like protein fusions or tags to measure their secretion. Another common approach to study secretion is to evaluate how a SS directs secretion of a reporter protein, for example a fluorescent protein (Barrero et al., 2018; Fischer & Requena, 2022; Gordon et al., 2000; Huang & Shusta, 2005; Lu et al., 2021). Many approaches aiming at studying secretion have developed systems that do not require a reporter protein, allowing the secretion of proteins of interest by using short detection tags (Cleaver et al., 2025; Karimian et al., 2025; O'Riordan et al., 2024; Püllmann et al., 2021). An interesting approach to study secretion in yeast is the implementation of Yeast Surface Display (YSD). YSD enables the presentation of heterologous proteins on the cell wall of yeast cells. This system relies on the fusion of a protein of interest to a yeast cell wall-anchoring domain which tethers the target protein to the cell surface. It is widely used for protein engineering, peptide screening, enzyme evolution, and can be applied for the study of protein secretion (Boder & Wittrup, 1997; Kishkevich et al., 2025; Linciano et al., 2025; McMahon et al., 2018; O'Riordan et al., 2024). However, because YSD requires the fusion of a cell wall-anchoring domain to the secreted protein, and many YSD systems rely on big anchor proteins, the structure and activity of the secreted protein can be compromised (Inokuma et al., 2020; Zhang et al., 2022). One common YSD approach employs the a-agglutinin system from *S. cerevisiae*. A-agglutinin is a cell-surface heterodimeric protein complex playing a crucial role during mating. It is composed of Aga1p, a 725 amino acid protein bound to the exterior of the cell wall via a glycosylphosphatidylinositol (GPI) anchor, and Aga2p, the second a-agglutinin subunit, which is an 87 amino acid protein including a SS that is subsequently cleaved. Aga1p and Aga2p interact by the formation of two disulfide bonds to form a-agglutinin on the cell surface (Cappellaro et al., 1991; Roy et al., 1991). This interaction is strong and can be used to display proteins on the yeast surface. If Aga2p is fused to a protein and this fusion is in contact with the cell surface where Aga1p is present, the Aga2p moiety will interact with Aga1p and the fused protein will be displayed on the surface of the cell wall (Boder & Wittrup, 1997; O'Riordan et al., 2024). Leveraging this mechanism, YSD has recently

been adapted as a high-throughput platform to quantitatively probe protein secretion efficiency in *S. cerevisiae* (Kishkevich et al., 2025).

In this study, all the potential SSs in *K. marxianus* DMKU3-1042 were identified *in silico* and classified based on three parameters affecting secretion: length, isoelectric point and hydrophobicity. A custom Design of Experiments (DoE) approach was applied to select a subset of 11 varied SSs. To test their secretion efficiency, a *K. marxianus* platform strain for YSD was constructed. The secretory efficiency of the 11 SSs was assessed with three different proteins: α-galactosidase Mel1p, Venus fluorescent protein and the granulocyte-macrophage colony-stimulating factor precursor Csf2p. Enzymatic, fluorescence and YSD assays allowed the characterisation of the SSs. The tools constructed here constitute a further expansion of the YTK toolkit for new applications related to protein secretion.

Materials and Methods

Strains, maintenance and preculture

All *K. marxianus* strains used in this study (Table S1) are derived from *K. marxianus* ASR.140 (WT), a modified strain to express B clade lactose transporter (Varela et al., 2019). This strain is in turn derived from *K. marxianus* NBRC 1777. Strains were cultured in YPD medium (yeast extract, 10 g L^{-1}; peptone, 20 g L^{-1}; dextrose, 20 g L^{-1}) for 20 – 24 hours and mixed with sterile 60% (v/v) glycerol to a final concentration of 24% (v/v) and 1 mL aliquots of this mixture were stocked frozen at -80 °C. For cultivation, a frozen stock vial was used to streak a YPD agar plate (YPD + agar 2%) that was incubated for 24 – 48 hours at 30 °C. After that, the plate was stored at 4 °C for up to 1 month. Inoculum was prepared by transferring a single colony from the YPD plate into sterile YPD. After 18 – 20 hours of growth on an orbital shaker 200 rpm at 30 °C, cells were directly inoculated to an initial Absorbance (A_{600}) of 0.1.

Genes to assess secretion

The CDS of *S. cerevisiae* α-galactosidase *MEL1* (NCBI: CUW00896.1) and *Homo sapiens* granulocyte-macrophage colony-stimulating factor precursor *CSF2* (NCBI: NP_000749.2) were downloaded and codon optimised using SnapGene software (https://www.snapgene.com/), removing BsaI and BsmbI internal sites. Upstream and Downstream GG sequences were attached as described by (Lee et al., 2015). The corresponding DNA sequences were synthetised by Twist Bioscience (South San Francisco, CA, USA). To develop a YSD-adapted Venus, the CDS coding for Venus fluorescent pYTK033 was used as PCR template, (MoClo Yeast Toolkit, #1000000061, Addgene) (Lee et al., 2015).

Selection of a set of native secretion signals

To select a diverse set of SSs, the proteome of *K. marxianus* strain DMKU3-1042 was retrieved from the UniProt database (UP000065495). Subsequently, the putative secretome of this strain was obtained by filtering its proteome with SignalP6.0 (Teufel et al., 2022). All the proteins containing a SS with a certainty higher than 50% were preselected. This list was then curated by TargetP2.0 (Armenteros et al., 2019) and

TMHMM 2.0 (Krogh et al., 2001) to obtain a final list of 162 putative extracellular secreted proteins. In order to select a subset to be tested, the 162 SSs were organised by length, isoelectric point (IP) and hydrophobicity calculated as the GRAVY index (Gasteiger et al., 2005; Kyte & Doolittle, 1982) since these 3 features have an impact in the secretion efficiency (Delic et al., 2013; Nielsen et al., 2019; Yarimizu et al., 2015). These values were normalised to facilitate display. A custom DoE approach was followed due to the values of the features following a non-uniform experimental range. A total of 11 SSs were selected, 3 representing central points (median length, IP and GRAVY) and 8 representing axial/fractional points representing extreme cases of SSs (Table 1, Fig. S1), allowing the assessment of a very varied pool of native signals.

Construction of expression cassettes

All the primers and plasmids used in this work are listed in Table S2 and Table S3 respectively. SSs were amplified by designing partially self-annealing primers and Q5 High-Fidelity Polymerase (New England Biolabs, Ipswich, UK). Amplification was confirmed by gel electrophoresis (2% agarose, TAE buffer) and, when the expected band was present, subsequent purification from Q5 reaction was performed using GeneJet PCR purification kit (Fisher Scientific) and eluted in water prior to use. The purified PCR products were cloned into the backbone plasmid pYTK001 following GG protocol (Lee et al., 2015) and competent *E. coli* DH5α cells were transformed as previously described (Rajkumar et al., 2019). Fragments comprising the *MEL1* and *CFS2* CDSs were cloned likewise into pYTK001. All plasmid sequences were confirmed by Sanger sequencing (Eurofins) and stocked in glycerol for future constructions.

Assembly of expression cassettes to assess secretion of Mel1p or Venus in yeast were performed following the YTK protocol (Lee et al., 2015) and using our *K. marxianus* toolkit (Rajkumar et al., 2019). Briefly, a GG reaction was prepared mixing the preassembled integrative backbone targeting I4 integration site in *K. marxianus* ASR.140 chromosome IV (ChrIV:240042..241741), *KmPDC1* promoter, pYTK045 (Venus part type 3b) or pnoSSMEL1noOH, *KmPDC1* terminator and one of the secretion signals to test. Assembly of expression cassettes to test YSD of Mel1p, Venus and Csf2p in yeast were performed likewise, substituting the previous CDS part for

pVenus_agasys (pYTK045 with adapted overhangs), pnoSSCSF2 or pScMEL1_agasys and adding the pAGA2-HA fusion tag.

Bacterial transformation

Bacterial transformation was followed as described by (Yang et al., 2022) and heat shock was performed on thermocycler (10 min 4 °C, 1 min 42 °C, 2 min 4 °C, and 1 h 37 °C). Cells were plated on LB agar (Formedium, Hunstanton, UK) supplemented with the required antibiotic.

Yeast transformation

K. marxianus was transformed as described (Gietz & Schiestl, 2007) with a few differences: 1 µg of integrative plasmid carrying the expression cassette was digested with SgsI (FD1894, Fisher Scientific) in a total volume of 35 µL and the digested plasmid was combined with transformation mix and *K. marxianus* cells prepared as per the protocol. Heat shock was carried for 25 minutes. Yeast cells were selected on YPD plates supplemented with 200 mg L^{-1} hygromycin (Fisher Scientific Dublin, Ireland). Successful integrations were confirmed by screening yeast colonies by PCR using two sets of primers: one amplifying the product between a region upstream of an integration site and the promoter of the integrated construct while the second set of primers amplified the product between a region downstream the integration site and the selective marker gene sequence.

Construction of *K. marxianus* YSD platform strain.

To construct a *K. marxianus* platform strain capable of displaying secreted proteins, *K. marxianus* ASR.140 was transformed with pRF3_TEF1pr to express *ScAGA1* for display of the fusion constructs via CRISPR/Cas9 as previously described (Rajkumar & Morrissey, 2022). Briefly, a gRNA targeting *KmAGA2* was generated and cloned in pUCC001 (pUCC-AGA2T2). Concurrently, a Repair Fragment (RF) was designed carrying *ScAGA1* under the control of *ScTEF1* promoter (pYTK013) and the KMXK_A03020t terminator (KmK.T3, Addgene Plasmid #125055) with the upstream and downstream regions of *KmAGA2* flanking the expression cassette. This RF was assembled into an *E. coli* plasmid and stocked (pRF3_TEF1pr, Fig. S2). Prior to transformation, pRF3_TEF1pr was digested with SgsI (FD1894, Fisher Scientific) and

30 µL of this digested product was used when transforming *K. marxianus* together with 4 µL of pUCC-AGA2T2.

α-galactosidase assay

To assess the activity of secreted α-galactosidase, a modified version of a previously published reporter assay was performed (Álvarez-Cao et al., 2019). To a 540 µL aliquot of supernatant, 60 µL of PNPG solution was added (100 mM p-nitrophenyl-α-d-galactopyranoside (PNPG, Sigma Aldrich), 610 mM citric acid and 770 mM Na_2HPO_4, pH 4) and this mix was vortexed and incubated at 40 °C for 1 hour. The reaction was stopped by adding 300 µL of 1 M Na_2CO_3 and the colour was measured at 400 nm. These values were normalised to population size dividing by A_{600}. Statistical significance was assessed via pairwise Tukey's Honestly Significant Difference (HSD) tests (alpha = 0.05), and results were grouped by letter codes to denote non-significant differences (Python statsmodels).

Growth assays on melibiose

In order to study the effect of different SSs on the secretion of functional α-galactosidase, *K. marxianus* strains were precultured on 5 mL YPD overnight from YPD plates. The following day, strains were inoculated in Minimal Media Melibiose (MMMel: 5 g L^{-1} $(NH_4)_2SO_4$, 3 g L^{-1} KH_2PO_4, 0.5 g L^{-1} $MgSO_4 \cdot 7\ H_2O$, 20 g L^{-1} Melibiose (Thermo scientific, Japan)) (Verduyn et al., 1992). Growth was followed in a Biolector Microbioreactor (Beckman Coulter, USA) at 30 °C, aeration 10 mL min^{-1} for 72 hours.

Measurement of fluorescence in supernatant

To measure the fluorescence remaining in supernatant of grown cultures, a set of strains to be tested was inoculated in 5 mL of YPD overnight from YPD plates. The following day, a 96-well microtiter plate was prepared containing 200 µL YPD per well. Each strain was inoculated in triplicate to a starting A_{600} of 0.5 and incubated at 30 °C with gentle agitation. After 6 hours, the plate was centrifuged (Allegra V-15R benchtop centrifuge, Beckman Coulter), and the supernatant was transferred to a new 96-well microtiter plate (265301 96-well Black/Clear Bottom Plate, Thermo Scientific Nunc, USA) to measure fluorescence. Fluorescence was measured using a ClarioSTAR plus (BMG labtech, Ortenberg, Germany) with excitation and emission set

to 495 and 540 nm, respectively (bandwidth 20 nm). After correcting for the autofluorescence of wild-type *K. marxianus*, the signal was normalised to population size dividing by A_{600} and expressed as Relative Fluorescence Units (RFU). Statistical significance was assessed via pairwise Tukey's HSD tests (alpha = 0.05), and results were grouped by letter codes to denote non-significant differences (Python statsmodels).

Yeast Surface Display detection

Flow cytometry was performed in the SUSFERM Research Infrastructure at UCC (https://www.ucc.ie/en/susferm/). For the detection of yeast cells displaying secreted proteins, test strains were inoculated in 5 mL of YPD overnight from YPD plates. The following day, a 96-well microtiter plate (83.3925.500, Sarstedt, Germany) was prepared containing 200 µL YPD per well. Each strain was inoculated in triplicate to a starting A_{600} of 0.1 and incubated at 30 °C with gentle agitation. After 3 hours, the plate was centrifuged, and the supernatant was discarded. Yeast cells were washed twice with PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 1.8 mM KH_2PO_4). After the second wash, yeast cells were incubated for one hour at room temperature with gentle agitation in PBS buffer supplemented with Alexa Fluor® 488 anti-HA.11 (Biolegend, San Diego, CA, USA) for the strains expressing Mel1p and Csf2p or Alexa Fluor® 647 anti-HA.11 (Biolegend, USA) for the strains expressing Venus following manufacturer specifications. After incubation, cells were centrifuged and washed once with PBS buffer and resuspended in 200 µL of PBS buffer. To detect immunolabelled cells, a CytoFLEX flow cytometer (Beckman Coulter, CA, USA) using a 488 nm laser together with a 525/30 bandpass filter or a 638 nm with a 660/10 bandpass filter was employed. The median RFUs of the yeast cell populations was used for comparisons. Statistical significance was assessed via pairwise Tukey's HSD tests (alpha = 0.05), and results were grouped by letter codes to denote non-significant differences (Python statsmodels).

Results

Development of an α-galactosidase-based tool to assess secretion signals

In order to evaluate the secretion efficiency of a SS in *K. marxianus*, a system that allows the interchangeability of SSs was devised. This was designed based on the YTK and uses the secreted enzyme α-galactosidase coded by *S. cerevisiae MEL1* as a read-out (Fig. 1).

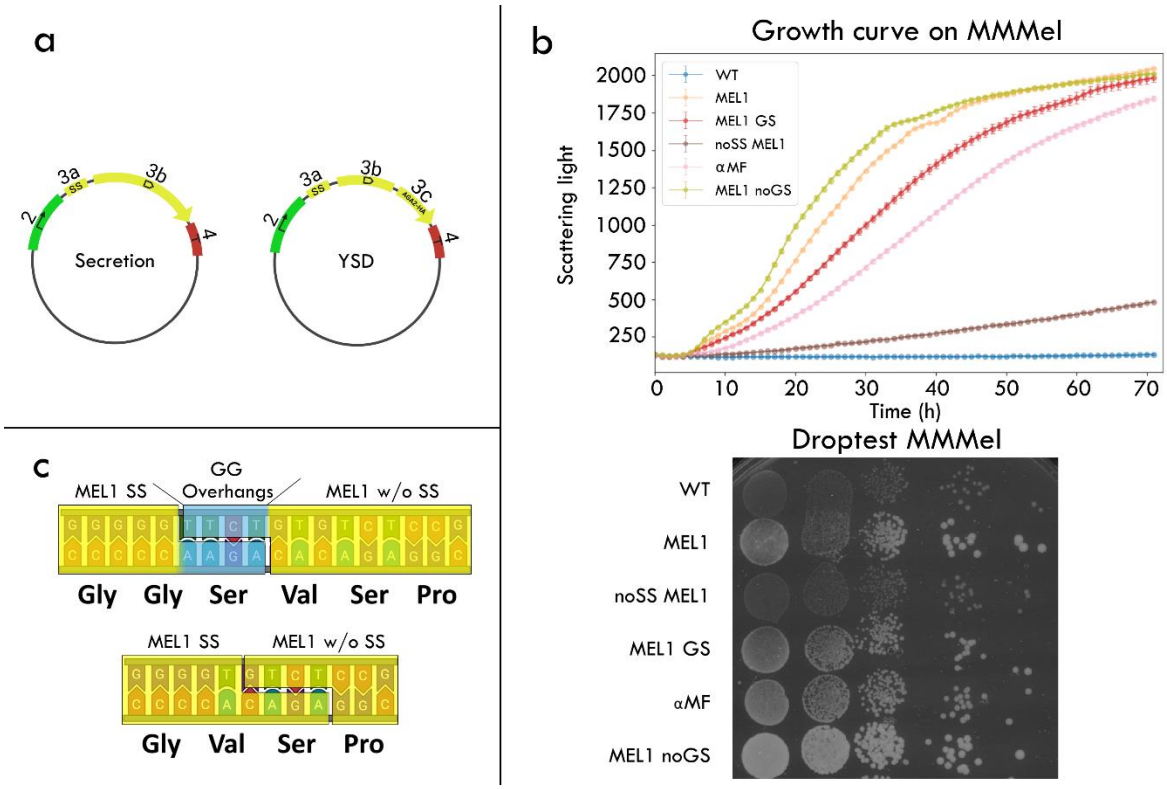


Fig. 1. a) Two types of YTK compatible plasmids were constructed and used in this work. Secretion plasmids were assembled with a promoter (YTK part type 2), SS (YTK part type 3a), coding sequence of reporter (YTK part type 3b) and terminator (YTK part type 4) in one GG reaction in an integrative expression vector. YSD plasmids follow similar approach but include an additional new YTK part type 3c that codes for the Aga2p and the HA tag for YSD and antibody labelling. This new part type 3c allows combination with any YTK part type 4. **b)** Growth on MMMel was used with control strains as a proof of concept: Wild type strain (ASR.140, blue), ASR.140 expressing *ScMEL1* (MEL1, orange), *ScMEL1* with SS removed (noSS MEL1, brown), *ScMEL1* fused to native *ScMEL1* SS with extra Gly – Ser from YTK (MEL1 GS, red). *ScMEL1* fused to αMF (αMF, pink). *ScMEL1* fused to native *ScMEL1* SS without extra Gly – Ser (MEL1 noGS, green). **c)** Original GG overhangs (TTCT) add two extra amino acids. Modified overhangs match the original sequence of *ScMEL1* (GTCT), avoiding the two extra linking amino acids and resulting in a seamless fusion.

Since the genome of *K. marxianus* does not encode an α-galactosidase or a melibiose transporter, if the SS attached to Mel1p does not direct secretion, there should be no α-galactosidase activity in the supernatant, and the yeast will not be able to grow on melibiose. Thus, by substituting different SS, the levels of secretion can be compared by simply growing the different yeasts on melibiose or by measuring enzymatic activity in the supernatant of liquid.

To assess secretion of α-galactosidase, a set of integrative expression plasmids was designed using YTK-based parts that would serve to construct different control strains (Fig. 1a). First, *MEL1* was cloned in the *K. marxianus* ASR.140 (in Fig. 1 named MEL1) to confirm this gene can confer growth on melibiose. Then, a second strain was constructed by removing the SS encoded in *MEL1* and cloning the non-secretable version (noSS MEL1). The native *MEL1* secretion signal was then turned into a YTK 3a part and fused to the version without signal, reconstituting the native gene and cloning this version again in ASR.140 (MEL1 GS). Finally, to test whether the system allows interchangeability of signals, the α Mating Factor (αMF) SS was also fused to the *MEL1* coding sequence lacking the signal and cloned in the chassis yeast (αMF in Fig. 1). The ability of these yeast strains to grow on MMMel was tested as well as the enzymatic activity present in the supernatant, confirming that *K. marxianus* ASR.140 (WT) is not able to grow on melibiose unless an α-galactosidase is expressed and secreted (Fig. 1b). The noSS MEL1 strain showed intracellular α-galactosidase activity (Fig. S3) and a slight growth on MMMel probably due to transmembrane diffusion of sugar. However, it was observed that strain MEL1 GS did not grow on MMMel at the same rate as MEL1. This was unexpected because, in principle, they were identical. Nevertheless, we noted the design of the YTK does not allow for a seamless fusion between SSs and the Mel1p reporter. The toolkit is designed to introduce two linking amino acids, Gly – Ser, when coding sequences are fused to tags. We hypothesised that, due to the brevity of SSs, the addition of these two extra amino acids could affect SS efficiency. A new construct was made to test this by modifying the YTK overhangs between the *MEL1* SS and the *MEL1* coding sequence to avoid the introduction of the Gly – Ser linkage. The new overhangs recreate the Val – Ser bond between the *MEL 1* SS and the *MEL1* coding sequence (Fig. 1c). This version of *MEL1* was cloned in

ASR.140 to generate another strain (MEL1 noGS) and this yeast was also assessed, this time recovering the same growth of the original *MEL1* signal (Fig.1b). To address the question of whether the introduction of the Gly – Ser bond causes an effect on secretion, or it was specific effect of the *MEL1* SS, the same analysis was performed with a set of nine other SSs (Fig. S4). Interestingly, in some cases, the presence of the extra amino acids improved secretion, while in most cases the effect was negative. These data confirm an impact of addition or not of extra amino acids at the junction of the SS and secreted protein.

Implementation of a Yeast Surface Display platform for *K. marxianus*

To study the secretion efficiency of SSs independently of the protein of interest, a second approach based on YSD and the ScAga1p/Aga2p system was adapted for *K. marxianus*. A strategy based on CRISPR/Cas9 was devised. First, we performed a homology search for ScAga1p and ScAga2p to ensure orthogonality of the system. We could not find an ScAga1p homolog in *K. marxianus* but a protein annotated as Aga2p with ~50% homology was found. We designed a CRISPR gRNA targeting *KmAGA2*. We also assembled a repair fragment containing *ScAGA1* under the control of *ScTEF1* promoter flanked by the upstream and downstream regions of *KmAGA2* (Fig. S2). Using our previously described CRISPR protocol (Rajkumar & Morrissey, 2022) we knocked out *KmAGA2* and substituted it with *ScAGA1*. We used this strain as a platform for YSD.

In the standard YTK architecture, the addition of a C-terminal fusion tag requires the use of the Type 4a part, which in turn demands a compatible Type 4b terminator (Fig. S5a). However, the vast majority of characterised terminators in existing MoClo toolkits are defined as Type 4 parts, rendering them incompatible with Type 4a tags. This creates a significant bottleneck for the combinatorial assembly of YSD constructs, where testing various terminators or tags is often required. To resolve this and enable high-throughput assembly using standard Type 4 terminators, we engineered a new architectural part: Type 3c (Fig. 1, Fig. S5b). This required the design of a modified coding sequence part, designated Type 3b**, which contains a unique 3' overhang to interface with the Type 3c tag (TGGC). This new Type 3c part serves as a bridge, accepting the Type 3b** CDS and outputting the standard overhang required for Type

4 terminators (ATCC). Consequently, this system allows for the seamless, combinatorial assembly of any SS (Type 3a), gene of interest (Type 3b**), and C-terminal tag (Type 3c) with the full library of standard Type 4 terminators.

To evaluate secretion in this platform, we assembled expression plasmids using the Yeast Toolkit (Lee et al., 2015) by fusing a SS to the N-terminal end of the protein to be secreted and a type 3c part containing the YSD tag composed of *ScAGA2* and the Hemagglutinin Antigen (HA) sequence (Fig. 1a). We tested this system by expressing ScMel1p fused to the αMF and measuring fluorescence in the supernatant of grown cultures (Fig. 2).

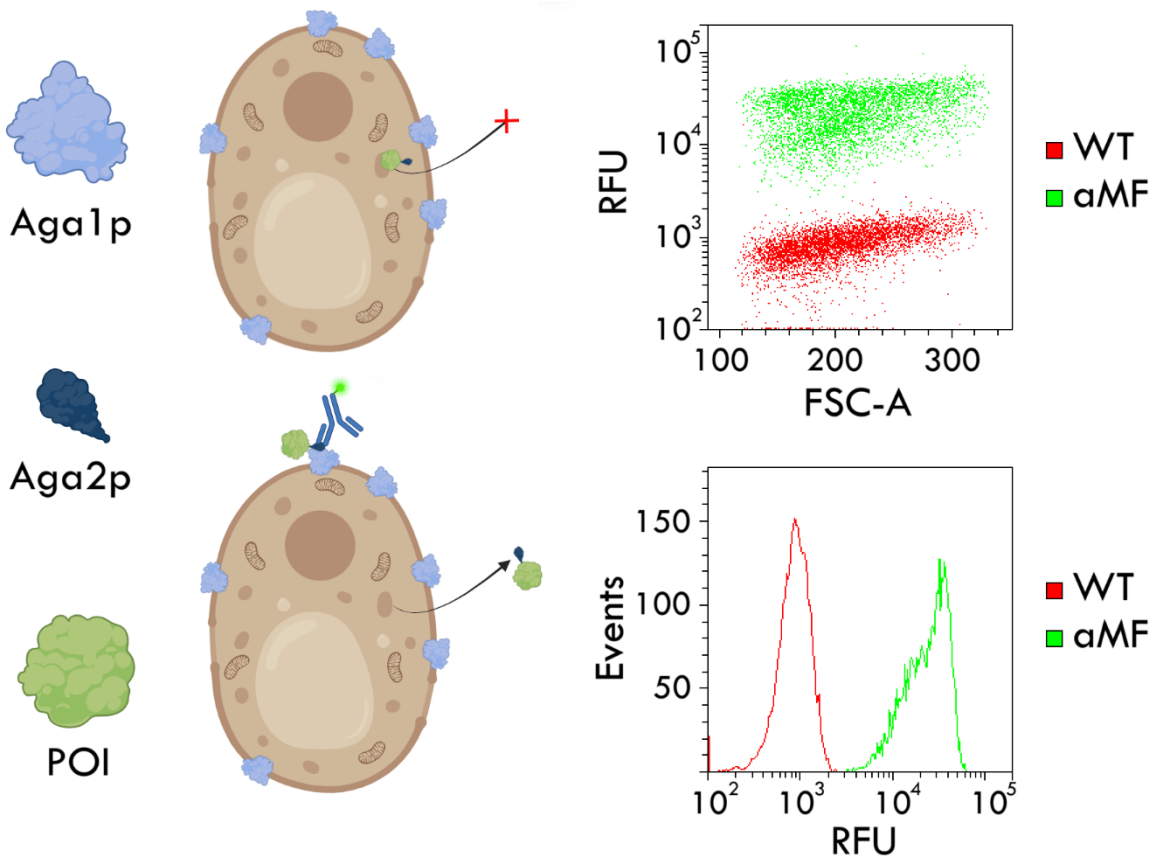


Fig. 2. Schematic view of the YSD system. The Aga1p is expressed and exposed on the outer part of the cell wall. Simultaneously, a fusion product composed of the Protein of Interest (POI), the Aga2p peptide and a small antigen is also synthetised by the cell. If the fusion product is correctly secreted, Aga2p interacts with Aga1p, displaying the POI and the antigen for antibody labelling. Labelled cells are detected by flow cytometry and compared. Strains expressing a functional secretion signal fused to a POI (αMF, green) can be easily identified from unlabelled cells (WT, red).

Identification of native secretion signals in *K. marxianus*

To select and test a pool of native SSs in *K. marxianus*, first the secretome of the strain DMKU3-1042 was identified. This yeast strain codes for 162 proteins with unique SSs (Table S4). These were classified based on their length, IP and GRAVY and displayed on a 3D space based on these 3 features (Figs. S1). We also performed a statistical analysis of the SSs for the three selected features (Fig. S6). In this yeast strain, the length of SSs ranges from 14 to 30 amino acids, although most common lengths go from 17 to 20 amino acids. Length of SSs does not follow a normal distribution. The IP of SSs ranges from 3.57 to 12.02 but they present a bi-modal distribution, with 5.27 and 8.5 being the most common. Hydrophobicity on the contrary, seems to be normally distributed along the range studied except for one outlier (W0TK22). By applying a custom DoE a diverse subset of 11 SSs was selected representing the most informative points of the DoE (Table 1). To confirm the diversity of these SSs a Principal Component Analysis (PCA) was also performed and it supported a great variability between the SSs selected (Fig. S6).

DoE points	Gene	UniProt	Sequence	Length	IP	GRAVY
Axial points	KLMX 30437	W0T8J1	MRFNRIAPFMLAPLASA	17	12.00	0.72
	ALG11	W0TFP9	MMSLAKFITFFLGLLLFLLVSLRVLS	26	11.00	2.00
	DSE4	W0T4U4	MLAQVWLLSLGYVLTTFLSAVGVAA	25	5.27	1.68
	ECM38	W0T4A4	MLFSTPYLWLLLLGSVDA	18	3.80	1.31
	GMC1	W0T2A3	MEFPNWWCVFGLLLAWLLSDTCEG	24	3.57	0.70
	GTB1	W0T6Y4	MLFPIAFLACVVCASG	16	5.27	2.23
	ROT1	W0T7W2	MRFLILLSIFAVAIRG	16	12.00	1.88
	SIM1 (SUN4)	W0TI90	MKFATALASASFIGATLVSALPHPHAKR	28	11.17	0.50
Central points	MSB2	W0TAK9	MQFSSIKVGLALLVLQSCAASVSA	24	7.98	1.31
	KLMX 10834	W0T5X0	MKSYYSLIGTILIGIINFAFA	21	8.25	1.34
	KRE1	W0TH48	MRNNVLVSLVALLAACAWVQ	20	8.00	1.35
Controls	MEL1	P04824	MFAFYFLTACISLKGVFG	18	7.95	1.56
	CSF2	P04141	MWLQSLLLLGTVACSI	16	5.27	1.68
	αMF*	P01149	MRFPSIFTAVLFAASSALAAPVNTTTED ETAQIPAEAVIGYSDLEGDFDVAVLPFS NSTNNGLLFINTTIASIAAKEEGVSLEKR EAEA	89	4.00	0.22

Table 1. List of SSs used in this study. A DoE approach was applied based on the Length, IP and GRAVY of each SS. * The αMF SS belongs to the post-translational secretion pathway and for this reason the DoE did not include it.

Assessment of secretion efficiency in *K. marxianus*

To assess the secretion efficiency of the collected SSs, three biotechnology-relevant proteins were selected: Mel1p, Venus fluorescent protein and the human granulocyte-macrophage colony-stimulating factor precursor (Csf2p). A set of *K. marxianus* strains was generated with integrated cassettes expressing these three proteins fused to the SS to be tested at the N-terminal end and the Aga2p fused to the HA epitope at the C-terminal end (Fig. 1a). The secretion efficiency of the SSs was measured for the three cases and compared with activity secretion and fluorescence secretion in the cases of Mel1p and Venus respectively.

We first assessed secretion of Mel1p by comparing how each SS promotes growth on MMMel (Fig. 3a). All the 11 SSs direct active Mel1p outside the cell except for Alg11p SS that showed no growth. The SSs of Klmx_30437p, Klmx_10834p, Rot1p and

Emc38p promote faster growth than the native Mel1p SS. Moreover, a commonly used SS, αMF, was one of the worst SSs, not reaching plateau by 48 hours of growth.

The level of secretion was then assessed by measuring the α-galactosidase activity present in the supernatant of grown cultures (Fig. 3b). Most of the SSs showed a similar relative level of activity compared to the growth assay (Fig. 3a) however slight differences were observed. We analysed these differences by applying a Tukey's HSD test, that categorised the SSs in groups with similar mean. This allowed the identification of 6 different groups based on their activity: group A show no secretion and includes WT, groups B and C include low-efficient SSs, group D that only includes Rot1p SS, group E that shows relatively good secretion and group F that includes the best SSs in this assay, including Mel1p.

We then applied the YSD platform to assess the efficiency of each SSs (Fig. 3c). Most of the SSs showed a similar relative level of secretion across the YSD and the α-galactosidase assays, with two clear exceptions: αMF and Sun4p SSs. This time, after applying Tukey's HSD test, up to 9 groups were identified, almost 1 per SS tested. Overall, we observed that all the SSs showed higher levels of secretion compared to the WT and in particular 3 SSs; Klmx_30437p, αMF and Sun4p, showed the highest level of secretion, above the native Mel1p. Following, Klmx_10834p, Ecm38p and Rot1p SSs showed similar levels to those of the native Mel1p. Finally, the remaining SSs lay below the levels of secretion detected for Mel1p SS.

We compared the results obtained with the different assays by means of a bubble plot (Fig. 3d). In general, we observed a similar efficiency of the SSs across the YSD and the activity assays, with two clear exceptions: αMF and Sun4p SSs. These two SSs, although showing the highest levels of secretion in the YSD assay, presented low α-galactosidase activity in both growth assay and reporter assay. This is interesting since αMF is the preferred SS when a heterologous protein is to be secreted. On the other hand, Klmx_30437p SS confirmed improved Mel1p secretion compared to the native Mel1p SS in both growth and enzymatic assays and Klmx_10834p, Ecm38p and Rot1p SSs showed similar levels to Mel1p SS.

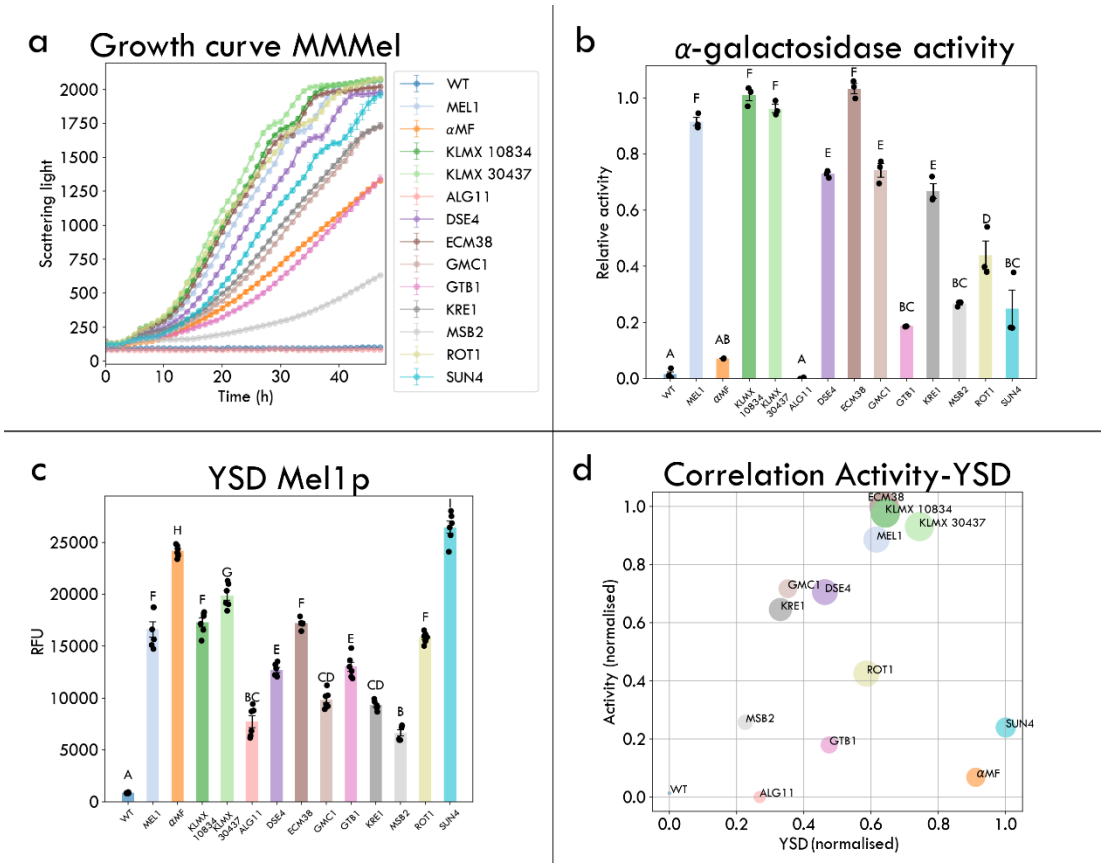


Fig. 3. a) Growth assay on MMMel of *K. marxianus* ASR.140 expressing ScMel1p fused to different SSs. WT (dark blue) and native ScMel1p SS (MEL1, light blue) are shown as controls. **b)** α-galactosidase activity detected in the supernatant of same set of *K. marxianus* strains. **c)** Detection of the secretion of ScMel1p with the same set of strains using the YSD system. **d)** Correlation between average α-galactosidase activity detected in supernatant (y-axis) and YSD (x-axis, normalised values) of cells expressing ScMel1p. Size of the bubbles is proportional to the μ_{max} on MMMel.

Similarly to Mel1p, a functional assessment of Venus secretion was performed by measuring fluorescence levels in the supernatant of yeast cells (Fig. 4a). In this case, many SSs did not show any significant fluorescence in the supernatant. This is the case for Klmx_30437p, Alg11p, Ecm38p, Gtb1p, Msb2p and Sun4p SSs, which all were grouped in the same category (A or AB) as the WT or Venus without SS attached. The rest of SSs showed an increase of fluorescence in the supernatant, with Dse4p and Klmx_10834p SSs being the best SSs in this assay followed by Rot1p, αMF, Kre1p and Gmc1p SSs. However, the variability during this assay did not allow the clear categorisation of many SSs in single groups. The secretion of Venus fluorescent

protein was also evaluated using the YSD platform (Fig. 4b). In this case, the SSs that did not show secretion of Venus (Klmx_30437p, Alg11p, Ecm38p, Gtb1p, Msb2p and Sun4p SSs) were able to display Venus through YSD at some level. Klmx_10834p and Dse4p SSs showed the highest levels of Venus display consistent as with the supernatant fluorescence assay. On the other hand, αMF, Kre1p, Rot1p and Gmc1p SSs showed similar levels of Venus display in the YSD, being classified in the EF group, but showed variable lower levels of fluorescence in the supernatant compared to Klmx_10834p and Dse4p SSs. This comparison is easier to visualise through a bubble graph (Fig. 4c) where we can observe that, while the YSD assay results in a scattered distribution of fluorescence levels, the measurement of fluorescence in the supernatant assay entails an all-or-nothing situation.

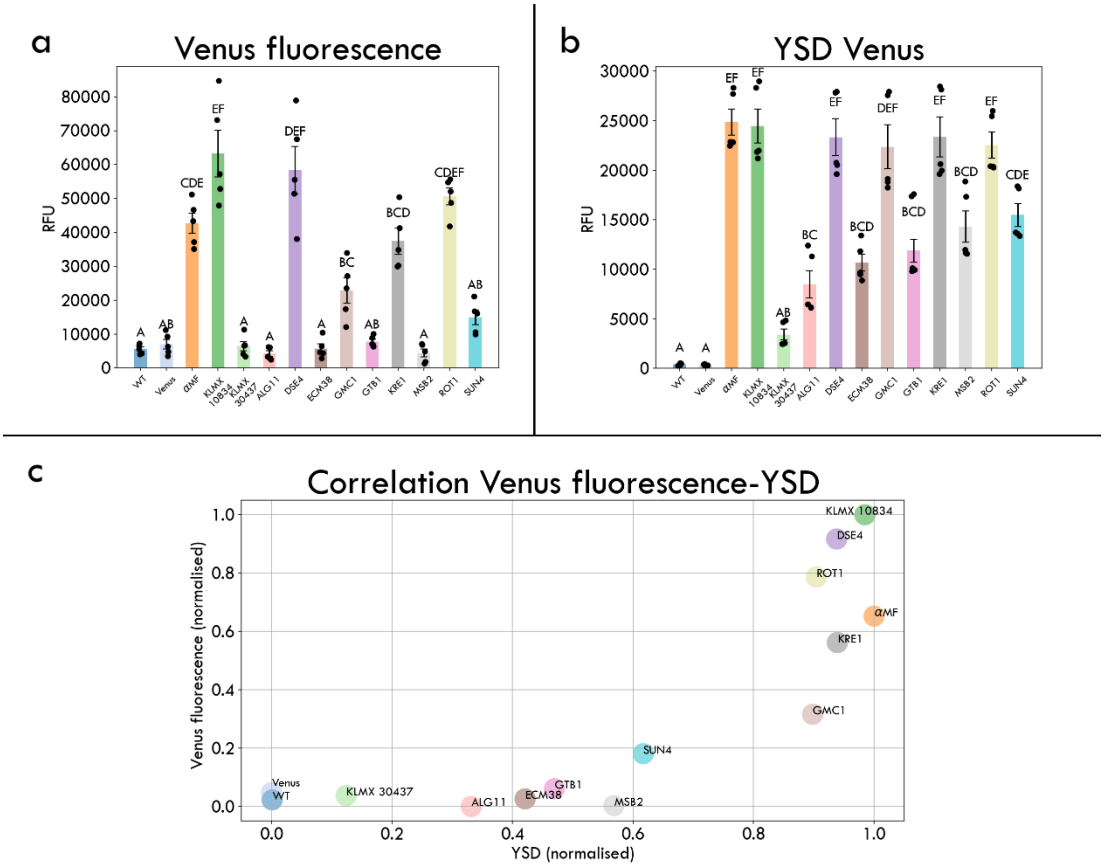


Fig. 4. a) Fluorescence detected in the supernatant of *K. marxianus* ASR.140 expressing Venus fused to different SSs. WT (dark blue) and WT expressing Venus without SS (Venus, light blue) are shown as controls. **b)** Detection of the secretion of Venus with the same set of strains using the YSD system. **c)** Correlation between average fluorescence detected in supernatant (y-axis) and YSD (x-axis, normalised values) of cells expressing Venus. Size of the bubbles is arbitrary.

Finally, the ability of the subset of 11 SSs to secrete a third protein, the human Csf2p was studied (Fig. 5). Many yeast SSs were able to secrete Csf2p more efficiently than the native Csf2p SS. For the secretion of Csf2p, the YSD results point to Dse4p, Klmx_10834p and αMF SSs in the HI group as the best signals, followed by Ecm38p, Kre1p and Gmc1p and finally on the F and E groups Gtb1p and Mbs2p SSs. All these SSs seem to be able to display more efficiently Csf2p than its native SS when expressed in *K. marxianus*. On the other hand, Rot1p SS was the only SS unable to display Csf2p, showing no statistical difference against the WT.

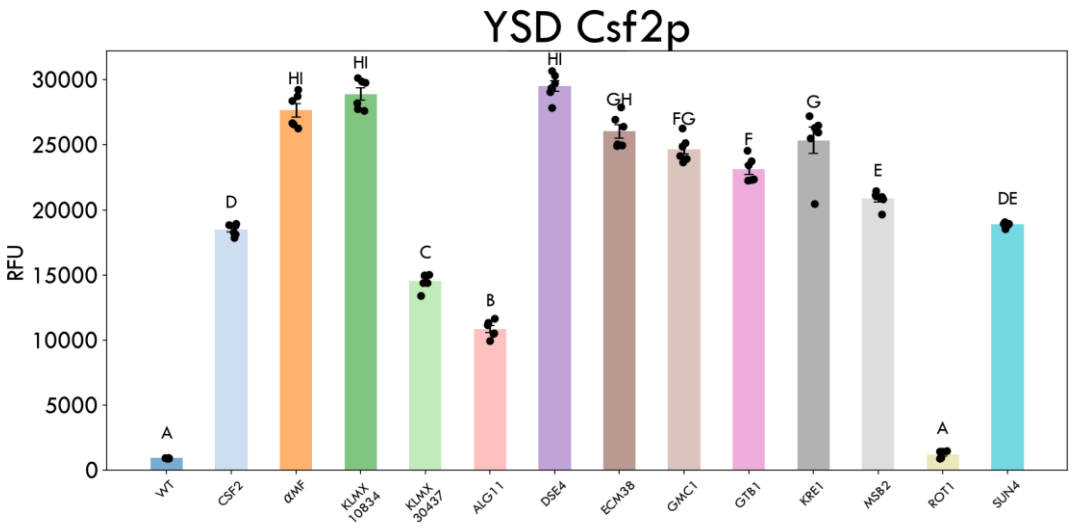


Fig. 5. Detection of the secretion of Granulocyte-macrophage colony-stimulating factor (Csf2p) directed by different SSSs using the YSD system. WT (dark blue) and WT expressing native HsCsf2p SS (CSF2, light blue) are shown as controls.

Assembly of a secretion toolkit expansion for *K. marxianus*

We conclude by presenting the secretion expansion for the *K. marxianus* toolkit (Table 2). The toolkit expansion contains (i) the characterised *K. marxianus* native SSs in the two forms developed in this work (with and without original YTK overhangs), (ii) different type 3 configurations for secreted or YSD assemblies, (iii) AGA2-HA as a new category part, type 3c, that allows fusions with modified-CDSs parts to YTK type 4 parts, enabling combinatorial assembly with terminators (Fig. S5), (iv) ScAGA1 to develop other YSD platforms and (v) two connectors that allow integration in a new

integrative site in *K. marxianus*. This type of part is compatible with YTK type 4 but not with original YTK type 3b and 4a parts.

Name	Part	Source	YTK part type	Description
KLMX_30437ss	SS1	*K. marxianus* DMKU3-1042	3a/SS	KLMX_30437 secretion signal.
ALG11ss	SS2	*K. marxianus* DMKU3-1042	3a/SS	ALG11 secretion signal.
DSE4ss	SS3	*K. marxianus* DMKU3-1042	3a/SS	DSE4 secretion signal.
ECM38ss	SS4	*K. marxianus* DMKU3-1042	3a/SS	ECM38 secretion signal.
GMC1ss	SS5	*K. marxianus* DMKU3-1042	3a/SS	GMC1 secretion signal.
GTB1ss	SS6	*K. marxianus* DMKU3-1042	3a/SS	GTB1 secretion signal.
MEL1ss	SS7	*K. marxianus* DMKU3-1042	3a/SS	MEL1 secretion signal.
MSB2ss	SS8	*K. marxianus* DMKU3-1042	3a/SS	MSB2 secretion signal.
ROT1ss	SS9	*K. marxianus* DMKU3-1042	3a/SS	ROT1 secretion signal.
SUN4ss	SS10	*K. marxianus* DMKU3-1042	3a/SS	SUN4 secretion signal.
10834noOHss	SS11	*K. marxianus* DMKU3-1042	3a*/SS	KLMX_10834 secretion signal that generates no Gly-Ser link.
30437noOHss	SS12	*K. marxianus* DMKU3-1042	3a*/SS	KLMX_30437 secretion signal that generates no Gly-Ser link.
ALG11noOHss	SS13	*K. marxianus* DMKU3-1042	3a*/SS	ALG11 secretion signal that generates no Gly-Ser link.
DSE4noOHss	SS14	*K. marxianus* DMKU3-1042	3a*/SS	DSE4 secretion signal that generates no Gly-Ser link.
ECM38noOHss	SS15	*K. marxianus* DMKU3-1042	3a*/SS	ECM38 secretion signal that generates no Gly-Ser link.
GMC1noOHss	SS16	*K. marxianus* DMKU3-1042	3a*/SS	GMC1 secretion signal that generates no Gly-Ser link.
GTB1noOHss	SS17	*K. marxianus* DMKU3-1042	3a*/SS	GTB1 secretion signal that generates no Gly-Ser link.
KRE1noOHss	SS18	*K. marxianus* DMKU3-1042	3a*/SS	KRE1 secretion signal that generates no Gly-Ser link.
MEL1noOHss	SS19	*K. marxianus* DMKU3-1042	3a*/SS	MEL1 secretion signal that generates no Gly-Ser link.
MSB2noOHss	SS20	*K. marxianus* DMKU3-1042	3a*/SS	MSB2 secretion signal that generates no Gly-Ser link.
ROT1noOHss	SS21	*K. marxianus* DMKU3-1042	3a*/SS	ROT1 secretion signal that generates no Gly-Ser link.
SUN4noOHss	SS22	*K. marxianus* DMKU3-1042	3a*/SS	SUN4 secretion signal that generates no Gly-Ser link.
noSSMEL1 3b	CDS1	*S. cerevisiae* NCYC 74	3b/CDS	ScMEL1 without secretion signal.
noSSMEL1 noOH	CDS2	*S. cerevisiae* NCYC 74	3b*/CDS	ScMEL1 without secretion signal. Modified DNA overhangs to avoid Gly-Ser link.
MEL1 Agasys	CDS3	*S. cerevisiae* NCYC 74	3b**/CDS	noSSMEL1 with modified DNA overhangs for YSD. To be used with type 3c**.
Venus Agasys	CDS4	Yeast Toolkit	3b**/CDS	Venus with modified DNA overhangs for YSD. To be used with type 3c**.
MEL1	CDS5	*S. cerevisiae* NCYC 74	3/CDS	ScMEL1 gene.

Name	Part	Source	YTK part type	Description
noSSMEL1	CDS6	*S. cerevisiae* NCYC 74	3/CDS	ScMEL1 gene without secretion signal.
ScAGA1	CDS7	*S. cerevisiae* S288C	3/CDS	ScAGA1 gene.
AGA2-HA	YSDt1	*S. cerevisiae* S288C	3c**/YSD tag	ScAGA2 fused to HA antigen adapted for YSD fusions. Not compatible with YTK 3b parts. Compatible with YTK 4 parts.
USRF3-AGA2	5'Hom	*K. marxianus* DMKU3-1042	1/5' Homology	Sequence upstream of *KmAGA2*. SgsI site for plasmid linearisation. Integration at Chromosome II:118 274..118 073 (in *KmAGA2*).
DSRF3-AGA2	3'Hom	*K. marxianus* DMKU3-1042	7/3' Homology	Sequence downstream of *KmAGA2*. SgsI site for plasmid linearisation. Integration at Chromosome II:118 274..118 073 (in *KmAGA2*).

Table 2. Contents of the *K. marxianus* secretion toolkit expansion. The expansion comprises 10 SSs with the original YTK DNA-overhangs and 12 SSs with modified YTK DNA-overhangs for seamless fusion with ScMEL1. The expansion also contains 5 different versions of the ScMEL1 CDS and a modified Venus CDS for YSD constructs. Finally, the expansion also contains a new part type, 3c, that encodes fusions that allow YSD (in our expansion ScAGA2 combined with the HA epitope). * Part that contain modified DNA overhangs for seamless fusion with ScMEL1 (CDS2). ** Part with YTK-modified DNA overhangs for YSD fusions.

The biological parts collected and characterised here are intended to provide a modular pathway for the assembly and assessment of secreted protein constructs by means of a reporter protein such as *ScMEL1* or Venus, or to be tested by YSD. While the parts can be used to construct *K. marxianus* expression systems in general, they are optimally compatible with the original YTK available from Addgene (#1000000061). GG assembly is an established in vitro assembly technique, and with this expansion we help to further establish its use as standard.

Discussion

Secretion is a desired feature when producing heterologous proteins because it greatly simplifies the purification, but its complexity and diversity amongst producing organisms makes it a difficult challenge to tackle. In this study, the identification and classification of 162 unique SSs of *K. marxianus* DMKU3-1042 (Table S4) provides a valuable foundation for investigating protein secretion in this industrially relevant yeast. By analysing three physicochemical properties, namely length, IP, and GRAVY, the SS landscape was effectively mapped for this yeast strain, revealing non-normal distributions in SS length and IP, and a near-normal distribution for hydrophobicity (Fig. 6S). Interestingly, the bimodal IP distribution, with two frequent values near pH 5.27 and 8.5, suggests the presence of two distinct SS classes that might serve different purposes. Moreover, the use of a DoE approach and PCA for selecting a representative subset of 11 SSs allowed statistical diversity, maximising information gain from a minimal SS set (Table 1). However, due to the diversity and complex predictability of secretion, more data would be needed to fill the landscape of SS efficiency. This classification based on SS features paves the way for more feature-driven studies of SSs in *K. marxianus* and, while the chosen features in this study are known to play a significant role in secretion efficiency, other features not explored here, such as secondary structure or post-translational modifications, might be more significant and add another layer of complexity (Nielsen et al., 2019). The advent of AI will certainly stimulate the development of AI-based pipelines for identification and classification of SSs, revolutionising the way we understand signal properties and functions.

MoClo is a modular strategy that fully aligns with the principles of synthetic biology and weaves seamlessly within the DBTL cycle. This explains the surge in ready-to-use toolkits for diverse organisms like the YTK in the last two decades. In this research, a secretion-focused expansion for the YTK was presented. This YTK secretion expansion tackles one of the few weaknesses of a broadly used MoClo toolkit: the alteration of SS properties by the introduction of the Gly – Ser linkage (Fig. S4). However, DNA overhangs are intrinsic to the MoClo methodology. This requirement constrains the turnover of the DBTL cycle because of the need to redesign the overhangs every time

a new protein is sought to be secreted. In this work the overhangs were tailored for Mel1p, but the SSs of the toolkit will add the first two amino acids of Mel1p to any other protein that is tested with this tool, altering its sequence and potentially interfering in the SS activity.

To test this MoClo secretion expansion and study the secretion efficiency of the SS subset, a YSD strategy was successfully adopted in *K. marxianus* by developing a platform strain capable of expressing and displaying secreted proteins (Fig. 2). To engineer *K. marxianus* into a YSD platform, the *ScAGA1* anchor was cloned using CRISPR Cas9 in substitution of *KmAGA2*. The development of this system for *K. marxianus* expands the biotechnological possibilities of this yeast with attractive industrial features like thermotolerance and rapid growth. By leveraging the *S. cerevisiae* Aga1p/Aga2p system, 3 different biotechnologically relevant proteins, the α-galactosidase Mel1p, the fluorescent protein Venus and the granulocyte-macrophage colony-stimulating factor precursor Csf2p, were selected to be combined with the SS subset and evaluate their secretion. Activity and fluorescence assays of Mel1p and Venus respectively were used to complement the YSD results (Figs. 3, 4). YSD is a powerful platform to study secretion, but some SSs, such as Sun4p and αMF SSs, showed great display efficiency while showing very poor enzymatic (Fig. 3b) or fluorescent (Fig. 4a) activity. We believe YSD introduces a bias that exaggerates secretion efficiency of SSs that are not fully secreted but kept embedded on the cell wall. The secretion of the third protein, Csf2p, was tested only through the YSD platform to further characterise the toolkit expansion (Fig. 5). Based on these results, it was demonstrated that the same SS can have different secretory impact depending on the protein to which is attached. All the SSs tested, except of αMF, belong to the co-translational pathway. Thus, they are identified and cleaved by the secretion machinery during early synthesis of the polypeptide (Delic et al., 2013). This SS processing reduces the impact the SS could have during folding of the rest of the polypeptide by unexpected interactions although, as observed, it does not prevent it (e.g. Klmx_30437p Figs. 3, 4). On the other hand, the αMF, a post-translational SS and one of the most common choices for heterologous secretion in yeasts, is particularly successful in the YSD assays but is a suboptimal option when activity assays are

employed to measure secretion. Due to its structure and length, 89 amino acids, it is possible that αMF SS interacts with the protein to which is fused during translation, preventing it from reaching its correct folding and therefore affecting the activity, but efficiently secreting and displaying it. This could be proved by assessing the secretion efficiency of other known post-translational SSs like Prc1p or Gas1p (Yi et al., 2021) by both methodologies (YSD and activity assays). Significantly, two native *K. marxianus* SSs, Klmx_10834p and Dse4p, showed superior secretion activity for all the targets and assays used, outperforming αMF and the native signals of the target proteins (Figs. 3, 4, 5). Additionally, Rot1p SS also showed high efficiency with ScMel1p and Venus but very poor performance when secreting Csf2p. This discrepancy could be further investigated. Moreover, by finding a functional Csf2p assay that tests the hormone's efficiency, a functional comparison such as those for ScMel1p and Venus (Figs. 3d, 4c) could be completed.

In this research, YSD was shown to be an efficient methodology to test secretion of proteins in a high-throughput fashion, however it was observed that YSD is not enough as a complete screening strategy and it should be coupled with functional assays. For some proteins this is challenging to do (e.g. structural proteins, transcription factors or chaperones) but it is crucial to find ways to directly measure their expected outcome (i.e. structure coherence, level of gene expression or unfolded protein level) in order to select optimal SSs (Kishkevich et al., 2025; Püllmann et al., 2021).

A challenge when study secretion while relying on MoClo-based DNA assembly resides on the need to introduce DNA overhangs that link parts. This overhangs, when present within translated parts will introduce undesired amino acids that, due to brevity of SSs will change the signals properties and interfere with their efficiency. Some approaches that do not rely on DNA overhangs could tackle this challenge, like intein-based systems (Gahyeon et al., 2025; Marshall et al., 2014). Furthermore, DNA assembly approaches are currently used because the cost of synthesising long DNA strands is still not competitive, however the moment gene-sized DNA strands reach competitive prices for high throughput applications, MoClo and similar modular strategies that aim at constructing long DNA strands from shorter pieces will become

pointless, speeding up the DBTL turnover and allowing easier assessment avoiding time-consuming *E. coli* cloning steps.

Another limitation of YSD is the need to attach labelling tags, or even full proteins, that display the secreted product on the cell wall. In the case of this study, Aga2p was chosen because of its limited size, i.e. 69 amino acids, although together with the antigen tag and the peptide linkers the full tag reaches up to 100 amino acids. These 100 amino acids are not negligible and will have an impact on protein folding and most possibly in its function. Other systems cleverly circumvent this issue with minimal or no tagging (Cleaver et al., 2025; Karimian et al., 2025) or by directly sequencing the expression cassette from highly performing strains (Kishkevich et al., 2025), avoiding this issue. However, amino acid nanopore sequencing together with simultaneous secretion assessment can offer high-throughput pooled screening of secretion mutants with no labelling or two-strain systems like the one used in this study (Brinkerhoff et al., 2021; Wei et al., 2023; Yusko et al., 2017). In any case, the development of this native SS toolkit and YSD platform gives *K. marxianus* access to more reliable secretion and assessment methodology, adding robustness to this yeast as a precision fermentation platform instead of relying on heterologous SSs.

Conclusion

In this work, the SS landscape of *K. marxianus* DMKU3-1042 was identified *in silico* and classified based on three parameters affecting secretion: length, isoelectric point and hydrophobicity. A custom DoE approach was applied to select a dedicated set of SSs from *K. marxianus* and a novel YSD platform was established in this thermotolerant yeast. A MoClo *K. marxianus* secretion expansion toolkit was gathered. The secretion expansion was assessed with 3 different proteins, a food enzyme, a fluorescent protein and a human hormone were tested and assessed by YSD, and functional assays. Results show that, although secretion is protein specific, some SSs show great potential for the secretion of different proteins. This secretion expansion enables modular and systematic evaluation of heterologous protein export, while the YSD platform provides a versatile assay to study protein secretion and display in a high-throughput and quantitative manner. Together, these resources constitute a valuable toolkit for both fundamental studies of secretion and the applied engineering of *K. marxianus* as a robust host for biotechnological protein production, accelerating synthetic biology applications in non-conventional yeasts.

References

Álvarez-Cao, M. E., Cerdán, M. E., González-Siso, M. I., & Becerra, M. (2019). Optimization of Saccharomyces cerevisiae α-galactosidase production and application in the degradation of raffinose family oligosaccharides. *Microbial Cell Factories*, *18*(1), 1–17. https://doi.org/10.1186/S12934-019-1222-X

An, J., Shang, N., Liu, W., Niu, Y., Liang, Q., Jiang, J., & Zheng, Y. (2024). A yeast surface display platform for screening of non-enzymatic protein secretion in Kluyveromyces lactis. *Applied Microbiology and Biotechnology*, *108*(1), 503. https://doi.org/10.1007/S00253-024-13342-9

Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., Von Heijne, G., Elofsson, A., & Nielsen, H. (2019). Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance*, *2*(5). https://doi.org/10.26508/LSA.201900429

Aza, P., Molpeceres, G., de Salas, F., & Camarero, S. (2021). Design of an improved universal signal peptide based on the a-factor mating secretion signal for enzyme production in yeast. *Cellular and Molecular Life Sciences*, *78*(7), 3691–3707. https://doi.org/10.1007/S00018-021-03793-Y

Balasubramanian, S., Chen, J., Wigneswaran, V., Bang-Berthelsen, C. H., & Jensen, P. R. (2021). Droplet-Based Microfluidic High Throughput Screening of Corynebacterium glutamicum for Efficient Heterologous Protein Production and Secretion. *Frontiers in Bioengineering and Biotechnology*, *9*, 668513. https://doi.org/10.3389/FBIOE.2021.668513

Baptista, M., & Domingues, L. (2022). Kluyveromyces marxianus as a microbial cell factory for lignocellulosic biomass valorisation. *Biotechnology Advances*, *60*, 108027. https://doi.org/10.1016/J.BIOTECHADV.2022.108027

Bassett, S., Suganda, J. C., & Da Silva, N. A. (2024). Engineering peroxisomal surface display for enhanced biosynthesis in the emerging yeast Kluyveromyces marxianus. *Metabolic Engineering*, *86*, 326–336. https://doi.org/10.1016/J.YMBEN.2024.10.014

Bhatwa, A., Wang, W., Hassan, Y. I., Abraham, N., Li, X. Z., & Zhou, T. (2021). Challenges Associated With the Formation of Recombinant Protein Inclusion Bodies in Escherichia coli and Strategies to Address Them for Industrial Applications. *Frontiers in Bioengineering and Biotechnology*, *9*, 630551. https://doi.org/10.3389/FBIOE.2021.630551

Biermann, L., Tadele, L. R., Henrique, E., Perino, B., Nicholson, R., Lilge, L., & Hausmann, R. (2025). Recombinant Production of Bovine αS1-Casein in Genome-Reduced Bacillus subtilis Strain IIG-Bs-20-5-1. *Microorganisms 2025, Vol. 13, Page 60*, *13*(1), 60. https://doi.org/10.3390/MICROORGANISMS13010060

Bilal, M., Ji, L., Xu, Y., Xu, S., Lin, Y., Iqbal, H. M. N., & Cheng, H. (2022). Bioprospecting Kluyveromyces marxianus as a Robust Host for Industrial Biotechnology. *Frontiers in Bioengineering and Biotechnology*, *10*, 851768. https://doi.org/10.3389/FBIOE.2022.851768

Boder, E. T., & Wittrup, K. D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology*, *15*(6), 553–557. https://doi.org/10.1038/NBT0697-553

Bonanni, D., Litrico, M., Ahmed, W., Morerio, P., Cazzorla, T., Spaccapaniccia, E., Cattani, F., Allegretti, M., Beccari, A. R., Del Bue, A., & Martin, F. (2023). A Deep Learning Approach to Optimize Recombinant Protein Production in Escherichia coli Fermentations. *Fermentation*, *9*(6), 503. https://doi.org/10.3390/FERMENTATION9060503

Brinkerhoff, H., Kang, A. S. W., Liu, J., Aksimentiev, A., & Dekker, C. (2021). Multiple rereads of single proteins at single–amino acid resolution using nanopores. *Science*, *374*(6574), 1509–1513. https://doi.org/10.1126/SCIENCE.ABL4381

Cappellaro, C., Hauser, K., Mrśa, V., Watzele, M., Watzele, G., Gruber, C., & Tanner, W. (1991). Saccharomyces cerevisiae a- and alpha-agglutinin: characterization of their molecular interaction. *The EMBO Journal*, *10*(13), 4081–4088. https://doi.org/10.1002/J.1460-2075.1991.TB04984.X

Caro-Astorga, J., Rogan, M., Malcı, K., Ming, H., Debenedictis, E., James, P., & Ellis, T. (2025). SubtiToolKit: a bioengineering kit for Bacillus subtilis and Gram-positive bacteria. *Trends in Biotechnology*, *43*(6), 1446–1469. https://doi.org/10.1016/J.TIBTECH.2025.02.004

Celińska, E., Borkowska, M., Białas, W., Korpys, P., & Nicaud, J.-M. (2018). Robust signal peptides for protein secretion in Yarrowia lipolytica: identification and characterization of novel secretory tags. *Applied Genetics and Molecular Biotechnology*, *102*(102), 5221. https://doi.org/10.1007/s00253-018-8966-9

Celińska, E., Ledesma-Amaro, R., Larroude, M., Rossignol, T., Pauthenier, C., & Nicaud, J. M. (2017). Golden Gate Assembly system dedicated to complex pathway manipulation in Yarrowia lipolytica. *Microbial Biotechnology*, *10*(2), 450–455. https://doi.org/10.1111/1751-7915.12605

Chaudhuri, B., Steube, K., & Stephan, C. (1992). The pro-region of the yeast prepro-α-factor is essential for membrane translocation of human insulin-like growth factor 1 in vivo. *European Journal of Biochemistry*, *206*(3), 793–800. https://doi.org/10.1111/J.1432-1033.1992.TB16986.X

Claes, K., Van Herpe, D., Vanluchene, R., Roels, C., Van Moer, B., Wyseure, E., Vandewalle, K., Eeckhaut, H., Yilmaz, S., Vanmarcke, S., Çıtak, E., Fijalkowska, D., Grootaert, H., Lonigro, C., Meuris, L., Michielsen, G., Naessens, J., van Schie, L., De Rycke, R., … Callewaert, N. (2024). OPENPichia: licence-free Komagataella phaffii chassis strains and toolkit for protein expression. *Nature Microbiology 2024 9:3*, *9*(3), 864–876. https://doi.org/10.1038/s41564-023-01574-w

Cleaver, A., Luo, R., Smith, O. B., Murphy, L., Schwessinger, B., & Brock, J. (2025). High-throughput optimisation of protein secretion in yeast via an engineered biosensor. *Trends in Biotechnology*, *43*(4), 838–867. https://doi.org/10.1016/J.TIBTECH.2024.11.010

Crozet, P., Navarro, F. J., Willmund, F., Mehrshahi, P., Bakowski, K., Lauersen, K. J., Pérez-Pérez, M. E., Auroy, P., Gorchs Rovira, A., Sauret-Gueto, S., Niemeyer, J., Spaniol, B., Theis, J., Trösch, R., Westrich, L. D., Vavitsas, K., Baier, T., Hübner, W., De Carpentier, F., … Lemaire, S. D. (2018). Birth of a Photosynthetic Chassis: A

MoClo Toolkit Enabling Synthetic Biology in the Microalga Chlamydomonas reinhardtii. *ACS Synthetic Biology*, *7*(9), 2074–2086. https://doi.org/10.1021/ACSSYNBIO.8B00251

Delic, M., Valli, M., Graf, A. B., Pfeffer, M., Mattanovich, D., & Gasser, B. (2013). The secretory pathway: exploring yeast diversity. *FEMS Microbiology Reviews*, *37*(6), 872–914. https://doi.org/10.1111/1574-6976.12020

Domenzain, I., Sánchez, B., Anton, M., Kerkhoven, E. J., Millán-Oropeza, A., Henry, C., Siewers, V., Morrissey, J. P., Sonnenschein, N., & Nielsen, J. (2022). Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0. *Nature Communications 2022 13:1*, *13*(1), 1–13. https://doi.org/10.1038/s41467-022-31421-1

Donzella, L., Varela, J. A., Sousa, M. J., & Morrissey, J. P. (2021). Identification of novel pentose transporters in *Kluyveromyces marxianus* using a new screening platform. *FEMS Yeast Research*, *21*(4), 26. https://doi.org/10.1093/femsyr/foab026

Dupuis, J. H., Cheung, L. K. Y., Newman, L., Dee, D. R., & Yada, R. Y. (2023). Precision cellular agriculture: The future role of recombinantly expressed protein as food. *Comprehensive Reviews in Food Science and Food Safety*, *22*(2), 882–912. https://doi.org/10.1111/1541-4337.13094

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE*, *3*(11), e3647. https://doi.org/10.1371/JOURNAL.PONE.0003647

Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D. G., Patron, N. J., & Marillonnet, S. (2014). A Golden Gate modular cloning toolbox for plants. *ACS Synthetic Biology*, *3*(11), 839–843. https://doi.org/10.1021/SB4001504

Fasim, A., More, V. S., & More, S. S. (2021). Large-scale production of enzymes for biotechnology uses. *Current Opinion in Biotechnology*, *69*, 68–76. https://doi.org/10.1016/J.COPBIO.2020.12.002

Fitzgerald, I., & Glick, B. S. (2014). Secretion of a foreign protein from budding yeasts is enhanced by cotranslational translocation and by suppression of vacuolar targeting. *Microbial Cell Factories*, *13*(1), 1–12. https://doi.org/10.1186/S12934-014-0125-0

Fonseca, J. P., Bonny, A. R., Kumar, G. R., Ng, A. H., Town, J., Wu, Q. C., Aslankoohi, E., Chen, S. Y., Dods, G., Harrigan, P., Osimiri, L. C., Kistler, A. L., & El-Samad, H. (2019). A Toolkit for Rapid Modular Construction of Biological Circuits in Mammalian Cells. *ACS Synthetic Biology*, *8*(11), 2593–2606. https://doi.org/10.1021/ACSSYNBIO.9B00322

Gahyeon, K., Dukwon, L., Ji Hun, K., Seong Do, K., Hongki, K., Jae Heon, K., Sung Sun, Y., Soo-Jin, Y., Jay D., K., & Byung-Kwan, C. (2025). Engineering modular enzyme assembly: synthetic interface strategies for natural products biosynthesis applications. *Natural Product Reports*, *42*, 1489–1506. https://doi.org/10.1039/D5NP00027K

Gao, Y., Feng, X., Xian, M., Wang, Q., & Zhao, G. (2013). Inducible cell lysis systems in microbial production of bio-based chemicals. *Applied Microbiology and Biotechnology*, *97*(16), 7121–7129. https://doi.org/10.1007/S00253-013-5100-X

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. *The Proteomics Protocols Handbook*, 571–607. https://doi.org/10.1385/1-59259-890-0:571

Ghag, S. B., Adki, V. S., Ganapathi, T. R., & Bapat, V. A. (2021). Plant Platforms for Efficient Heterologous Protein Production. *Biotechnology and Bioprocess Engineering 2021 26:4*, *26*(4), 546–567. https://doi.org/10.1007/S12257-020-0374-1

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods 2009 6:5*, *6*(5), 343–345. https://doi.org/10.1038/nmeth.1318

Gietz, R. D., & Schiestl, R. H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, *2*(1), 31–34. https://doi.org/10.1038/nprot.2007.13

Gombert, A. K., Madeira, J. V., Cerdán, M. E., & González-Siso, M. I. (2016). Kluyveromyces marxianus as a host for heterologous protein synthesis. *Applied Microbiology and Biotechnology*, *100*(14), 6193–6208. https://doi.org/10.1007/s00253-016-7645-y

Gomes, A. M. V., Carmo, T. S., Carvalho, L. S., Bahia, F. M., & Parachin, N. S. (2018). Comparison of Yeasts as Hosts for Recombinant Protein Production. *Microorganisms*, *6*(2), 38. https://doi.org/10.3390/MICROORGANISMS6020038

Hagman, A., Säll, T., & Piškur, J. (2014). Analysis of the yeast short-term Crabtree effect and its origin. *FEBS Journal*, *281*(21), 4805–4814. https://doi.org/10.1111/febs.13019

Hong, M., Li, T., Xue, W., Zhang, S., Cui, L., Wang, H., Zhang, Y., Zhou, L., Gu, Y., Xia, N., & Li, S. (2022). Genetic engineering of baculovirus-insect cell system to improve protein production. *Frontiers in Bioengineering and Biotechnology*, *10*, 994743. https://doi.org/10.3389/FBIOE.2022.994743

Idiris, A., Tohda, H., Kumagai, H., & Takegawa, K. (2010). Engineering of protein secretion in yeast: strategies and impact on protein production. *Applied Microbiology and Biotechnology 2010 86:2*, *86*(2), 403–417. https://doi.org/10.1007/S00253-010-2447-0

Inokuma, K., Kurono, H., den Haan, R., van Zyl, W. H., Hasunuma, T., & Kondo, A. (2020). Novel strategy for anchorage position control of GPI-attached proteins in the yeast cell wall using different GPI-anchoring domains. *Metabolic Engineering*, *57*, 110–117. https://doi.org/10.1016/J.YMBEN.2019.11.004

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., & Liew, K. Bin. (2024). Evolving Paradigms of Recombinant Protein Production in Pharmaceutical Industry: A

Rigorous Review. *Sci 2024, Vol. 6, Page 9*, *6*(1), 9. https://doi.org/10.3390/SCI6010009

Karimian, P., Johnston, E., Kasprzak, A., Liu, J. W., Stockwell, S., Vanhercke, T., & Hartley, C. J. (2025). Use of a dual biosensor for identification of novel secretion signal peptides and efficient screening of precision fermentation production strains. *Future Foods*, *11*, 100640. https://doi.org/10.1016/J.FUFO.2025.100640

Kishkevich, A., Ciurkot, K., & Ellis, T. (2025). YTK Display-and-Secrete: Screening for Optimal Protein Secretion Elements in Saccharomyces cerevisiae. *ACS Synthetic Biology*. https://doi.org/10.1021/ACSSYNBIO.5C00264

Kittler, S., Müller, F., Elshazly, M., Wandrey, G. B., Klein, T., Daub, A., Spadiut, O., & Kopp, J. (2025). Transferability of bioprocessing modes for recombinant protease production: from fed-batch to continuous cultivation with Bacillus licheniformis. *BMC Biotechnology*, *25*(1), 1–13. https://doi.org/10.1186/S12896-025-00947-9

Knoshaug, E. P., Vidgren, V., Magalhães, F., Jarvis, E. E., Franden, M. A., Zhang, M., & Singh, A. (2015). Novel transporters from Kluyveromyces marxianus and Pichia guilliermondii expressed in Saccharomyces cerevisiae enable growth on l-arabinose and d-xylose. *Yeast*, *32*(10), 615–628. https://doi.org/10.1002/yea.3084

Köppl, C., Lingg, N., Fischer, A., Krößß, C., Loibl, J., Buchinger, W., Schneider, R., Jungbauer, A., Striedner, G., & Cserjan-Puschmann, M. (2022). Fusion Tag Design Influences Soluble Recombinant Protein Production in Escherichia coli. *International Journal of Molecular Sciences*, *23*(14), 7678. https://doi.org/10.3390/IJMS23147678

Kouprina, N., & Larionov, V. (2006). TAR cloning: insights into gene function, long-range haplotypes and genome structure and evolution. *Nature Reviews Genetics 2006 7:10*, *7*(10), 805–812. https://doi.org/10.1038/nrg1943

Krogh, A., Larsson, B., Von Heijne, G., & Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden markov model: application to

complete genomes. *Journal of Molecular Biology*, *305*(3), 567–580. https://doi.org/10.1006/JMBI.2000.4315

Kumar, P., Sahoo, D. K., & Sharma, D. (2021). The identification of novel promoters and terminators for protein expression and metabolic engineering applications in Kluyveromyces marxianus. *Metabolic Engineering Communications*, *12*, e00160. https://doi.org/10.1016/J.MEC.2020.E00160

Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology*, *157*(1), 105–132. https://doi.org/10.1016/0022-2836(82)90515-0

Lane, M. M., & Morrissey, J. P. (2010). Kluyveromyces marxianus: A yeast emerging from its sister's shadow. In *Fungal Biology Reviews* (Vol. 24, Issues 1–2, pp. 17–26). Elsevier. https://doi.org/10.1016/j.fbr.2010.01.001

Lee, M. E., DeLoache, W. C., Cervantes, B., & Dueber, J. E. (2015). A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology*, *4*(9), 975–986. https://doi.org/10.1021/sb500366v

Li, F., Chen, Y., Qi, Q., Wang, Y., Yuan, L., Huang, M., Elsemman, I. E., Feizi, A., Kerkhoven, E. J., & Nielsen, J. (2022). Improving recombinant protein production by yeast through genome-scale modeling using proteome constraints. *Nature Communications*, *13*. https://doi.org/10.1038/s41467-022-30689-7

Li, P., Fu, X., Zhang, L., Zhang, Z., Li, J., & Li, S. (2017). The transcription factors Hsf1 and Msn2 of thermotolerant Kluyveromyces marxianus promote cell growth and ethanol fermentation of Saccharomyces cerevisiae at high temperatures. *Biotechnology for Biofuels*, *10*(1), 289. https://doi.org/10.1186/s13068-017-0984-9

Linciano, S., Mazzocato, Y., Romanyuk, Z., Vascon, F., Farrera-Soler, L., Will, E., Xing, Y., Chen, S., Kumada, Y., Simeoni, M., Scarso, A., Cendron, L., Heinis, C., & Angelini, A. (2025). Screening macrocyclic peptide libraries by yeast display allows control of selection process and affinity ranking. *Nature Communications 2025 16:1*, *16*(1), 1–23. https://doi.org/10.1038/s41467-025-60907-x

Linder, T. (2019). Nitrogen assimilation pathways in budding yeasts. In *Non-conventional Yeasts: from Basic Research to Application* (pp. 197–236). Springer International Publishing. https://doi.org/10.1007/978-3-030-21110-3_7

Liu, D., Garrigues, S., & de Vries, R. P. (2023). Heterologous protein production in filamentous fungi. *Applied Microbiology and Biotechnology 2023 107:16*, *107*(16), 5019–5033. https://doi.org/10.1007/S00253-023-12660-8

Madhavan, A., Arun, K. B., Sindhu, R., Krishnamoorthy, J., Reshmy, R., Sirohi, R., Pugazhendi, A., Awasthi, M. K., Szakacs, G., & Binod, P. (2021). Customized yeast cell factories for biopharmaceuticals: from cell engineering to process scale up. *Microbial Cell Factories 2021 20:1*, *20*(1), 1–17. https://doi.org/10.1186/S12934-021-01617-Z

Marshall, C. J., Grosskopf, V. A., Moehling, T. J., Tillotson, B. J., Wiepz, G. J., Abbott, N. L., Raines, R. T., & Shusta, E. V. (2014). An Evolved Mxe GyrA Intein for Enhanced Production of Fusion Proteins. *ACS Chemical Biology*, *10*(2), 527–538. https://doi.org/10.1021/CB500689G

McMahon, C., Baier, A. S., Pascolutti, R., Wegrecki, M., Zheng, S., Ong, J. X., Erlandson, S. C., Hilger, D., Rasmussen, S. G. F., Ring, A. M., Manglik, A., & Kruse, A. C. (2018). Yeast surface display platform for rapid discovery of conformationally selective nanobodies. *Nature Structural & Molecular Biology 2018 25:3*, *25*(3), 289–296. https://doi.org/10.1038/s41594-018-0028-6

Nadar, C. G., Fletcher, A., Moreira, B. R. de A., Hine, D., & Yadav, S. (2024). Waste to protein: A systematic review of a century of advancement in microbial fermentation of agro-industrial byproducts. *Comprehensive Reviews in Food Science and Food Safety*, *23*(4), e13375. https://doi.org/10.1111/1541-4337.13375

Nielsen, H., Petsalaki, E. I., Zhao, L., & Stühler, K. (2019). Predicting eukaryotic protein secretion without signals. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, *1867*(12), 140174. https://doi.org/10.1016/J.BBAPAP.2018.11.011

Ntana, F., Mortensen, U. H., Sarazin, C., & Figge, R. (2020). Aspergillus: A Powerful Protein Production Platform. *Catalysts 2020, Vol. 10, Page 1064*, *10*(9), 1064. https://doi.org/10.3390/CATAL10091064

Nyathi, Y., Wilkinson, B. M., & Pool, M. R. (2013). Co-translational targeting and translocation of proteins to the endoplasmic reticulum. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1833*(11), 2392–2402. https://doi.org/10.1016/J.BBAMCR.2013.02.021

Obst, U., Lu, T. K., & Sieber, V. (2017). *A Modular Toolkit for Generating Pichia pastoris Secretion Libraries*. https://doi.org/10.1021/acssynbio.6b00337

O'Riordan, N. M., Jurić, V., O'Neill, S. K., Roche, A. P., & Young, P. W. (2024). A Yeast Modular Cloning (MoClo) Toolkit Expansion for Optimization of Heterologous Protein Secretion and Surface Display in Saccharomyces cerevisiae. *ACS Synthetic Biology*, *13*(4), 1246–1258. https://doi.org/10.1021/ACSSYNBIO.3C00743

Otto, M., Skrekas, C., Gossing, M., Gustafsson, J., Siewers, V., & David, F. (2021). Expansion of the Yeast Modular Cloning Toolkit for CRISPR-Based Applications, Genomic Integrations and Combinatorial Libraries. *ACS Synthetic Biology*, *10*(12), 3461–3474. https://doi.org/10.1021/ACSSYNBIO.1C00408

Owji, H., Nezafat, N., Negahdaripour, M., Hajiebrahimi, A., & Ghasemi, Y. (2018). A comprehensive review of signal peptides: Structure, roles, and applications. *European Journal of Cell Biology*, *97*(6), 422–441. https://doi.org/10.1016/J.EJCB.2018.06.003

Püllmann, P., Knorrscheidt, A., Münch, J., Palme, P. R., Hoehenwarter, W., Marillonnet, S., Alcalde, M., Westermann, B., & Weissenborn, M. J. (2021). A modular two yeast species secretion system for the production and preparative application of unspecific peroxygenases. *Communications Biology 2021 4:1*, *4*(1), 1–20. https://doi.org/10.1038/s42003-021-02076-3

Rajkumar, A. S., & Morrissey, J. P. (2022). Protocols for marker-free gene knock-out and knock-down in Kluyveromyces marxianus using CRISPR/Cas9. *FEMS Yeast Research*, *22*(1), 67. https://doi.org/10.1093/FEMSYR/FOAB067

Rajkumar, A. S., Varela, J. A., Juergens, H., Daran, J. M. G., & Morrissey, J. P. (2019). Biological parts for Kluyveromyces marxianus synthetic biology. *Frontiers in Bioengineering and Biotechnology*, *7*(MAY). https://doi.org/10.3389/fbioe.2019.00097

Rodrigues, F., Ludovico, P., & Leão, C. (2006). Sugar Metabolism in Yeasts: an Overview of Aerobic and Anaerobic Glucose Catabolism. In *Biodiversity and Ecophysiology of Yeasts*. https://doi.org/10.1007/3-540-30985-3_6

Roy, A., Lu, C. F., Marykwas, D. L., Lipke, P. N., & Kurjanh, J. (1991). The AGA1 Product Is Involved in Cell Surface Attachment of the Saccharomyces cerevisiae Cell Adhesion Glycoprotein a-Agglutinin. *Molecular and Cellular Biology*, *11*(8), 4196–4206. https://doi.org/10.1128/MCB.11.8.4196-4206.1991

Salazar, O., & Asenjo, J. A. (2007). Enzymatic lysis of microbial cells. *Biotechnology Letters*, *29*(7), 985–994. https://doi.org/10.1007/S10529-007-9345-2

Sarkari, P., Marx, H., Blumhoff, M. L., Mattanovich, D., Sauer, M., & Steiger, M. G. (2017). An efficient tool for metabolic pathway construction and gene integration for Aspergillus niger. *Bioresource Technology*, *245*, 1327–1333. https://doi.org/10.1016/J.BIORTECH.2017.05.004

Shaw, W. M., Khalil, A. S., & Ellis, T. (2023). A Multiplex MoClo Toolkit for Extensive and Flexible Engineering of Saccharomyces cerevisiae. *ACS Synthetic Biology*, *12*(11), 3393–3405. https://doi.org/10.1021/ACSSYNBIO.3C00423

Szybalski, W., Kim, S. C., Hasan, N., & Podhajska, A. J. (1991). Class-IIS restriction enzymes — a review. *Gene*, *100*(C), 13–26. https://doi.org/10.1016/0378-1119(91)90345-C

Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., Gíslason, M. H., Pihl, S. I., Tsirigos, K. D., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nature Biotechnology 2022 40:7*, *40*(7), 1023–1025. https://doi.org/10.1038/s41587-021-01156-3

Tong, Y., Jørgensen, T. S., Whitford, C. M., Weber, T., & Lee, S. Y. (2021). A versatile genetic engineering toolkit for E. coli based on CRISPR-prime editing. *Nature Communications 2021 12:1*, *12*(1), 1–11. https://doi.org/10.1038/s41467-021-25541-3

Varela, J. A., Puricelli, M., Ortiz-Merino, R. A., Giacomobono, R., Braun-Galleani, S., Wolfe, K. H., & Morrissey, J. P. (2019). Origin of Lactose Fermentation in Kluyveromyces lactis by Interspecies Transfer of a Neo-functionalized Gene Cluster during Domestication. *Current Biology*, *29*(24), 4284-4290.e2. https://doi.org/10.1016/j.cub.2019.10.044

Verduyn, C., Postma, E., Scheffers, W. A., & Van Dijken, J. P. (1992). Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, *8*(7), 501–517. https://doi.org/10.1002/YEA.320080703

Walhout, A. J. M., Temple, G. F., Brasch, M. A., Hartley, J. L., Lorson, M. A., Van den Heuvel, S., & Vidal, M. (2000). GATEWAY recombinational cloning: Application to the cloning of large numbers of open reading frames or ORFeomes. *Methods in Enzymology*, *328*, 575-IN7. https://doi.org/10.1016/S0076-6879(00)28419-X

Wei, X., Penkauskas, T., Reiner, J. E., Kennard, C., Uline, M. J., Wang, Q., Li, S., Aksimentiev, A., Robertson, J. W. F., & Liu, C. (2023). Engineering Biological Nanopore Approaches towards Protein Sequencing. *ACS Nano*, *17*(17), 16369. https://doi.org/10.1021/ACSNANO.3C05628

Xue, S., Liu, X., Pan, Y., Xiao, C., Feng, Y., Zheng, L., Zhao, M., & Huang, M. (2023). Comprehensive Analysis of Signal Peptides in Saccharomyces cerevisiae Reveals Features for Efficient Secretion. *Advanced Science*, *10*(2), 2203433. https://doi.org/10.1002/ADVS.202203433

Yang, Y., Liu, M., Wang, T., Wang, Q., Liu, H., Xun, L., & Xia, Y. (2022). An Optimized Transformation Protocol for Escherichia coli BW3KD with Supreme DNA Assembly Efficiency. *Microbiology Spectrum*, *10*(6). https://doi.org/10.1128/SPECTRUM.02497-22

Yarimizu, T., Nakamura, M., Hoshida, H., & Akada, R. (2015). Synthetic signal sequences that enable efficient secretory protein production in the yeast Kluyveromyces marxianus. *Microbial Cell Factories*, *14*(1), 1–14. https://doi.org/10.1186/S12934-015-0203-Y

Yi, J. K., Fujita, H., Mandon, E. C., & Gilmore, R. (2021). Rapid inactivation of the yeast Sec complex selectively blocks transport of post-translationally translocated proteins. *Journal of Biological Chemistry*, *297*(4), 101171. https://doi.org/10.1016/J.JBC.2021.101171

Yusko, E. C., Bruhn, B. R., Eggenberger, O. M., Houghtaling, J., Rollings, R. C., Walsh, N. C., Nandivada, S., Pindrus, M., Hall, A. R., Sept, D., Li, J., Kalonia, D. S., & Mayer, M. (2017). Real-time shape approximation and fingerprinting of single proteins using a nanopore. *Nature Nanotechnology*, *12*(4), 360–367. https://doi.org/10.1038/NNANO.2016.267

Zahrl, R. J., Prielhofer, R., Ata, Ö., Baumann, K., Mattanovich, D., & Gasser, B. (2022). Pushing and pulling proteins into the yeast secretory pathway enhances recombinant protein secretion. *Metabolic Engineering*, *74*, 36–48. https://doi.org/10.1016/J.YMBEN.2022.08.010

Zhang, C., Chen, H., Zhu, Y., Zhang, Y., Li, X., & Wang, F. (2022). Saccharomyces cerevisiae cell surface display technology: Strategies for improvement and applications. *Frontiers in Bioengineering and Biotechnology*, *10*, 1056804. https://doi.org/10.3389/FBIOE.2022.1056804

Ziegler, M., Zieringer, J., Döring, C. L., Paul, L., Schaal, C., & Takors, R. (2021). Engineering of a robust Escherichia coli chassis and exploitation for large-scale production processes. *Metabolic Engineering*, *67*, 75–87. https://doi.org/10.1016/J.YMBEN.2021.05.011

Zsebo, K. M., Lu, H. S., Fieschko, J. C., Goldstein, L., Davis, J., Duker, K., Suggs, S. V., Lai, P. H., & Bitter, G. A. (1986). Protein secretion from Saccharomyces cerevisiae directed by the prepro-alpha-factor leader region. *Journal of Biological Chemistry*, *261*(13), 5858–5865. https://doi.org/10.1016/S0021-9258(17)38462-4

Supplementary files

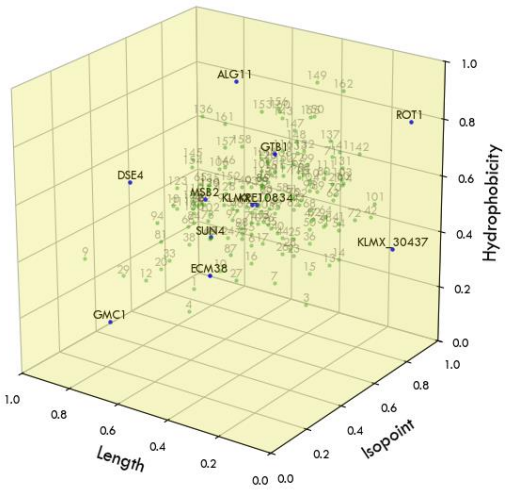


Fig. S1. 3-dimensional representation of Secretion Signals (SSs) based on three of their features normalised: Length, Isoelectric point (Isopoint), Hydrophobicity. In blue, disposition of the SSs selected in this study based on a custom DoE.

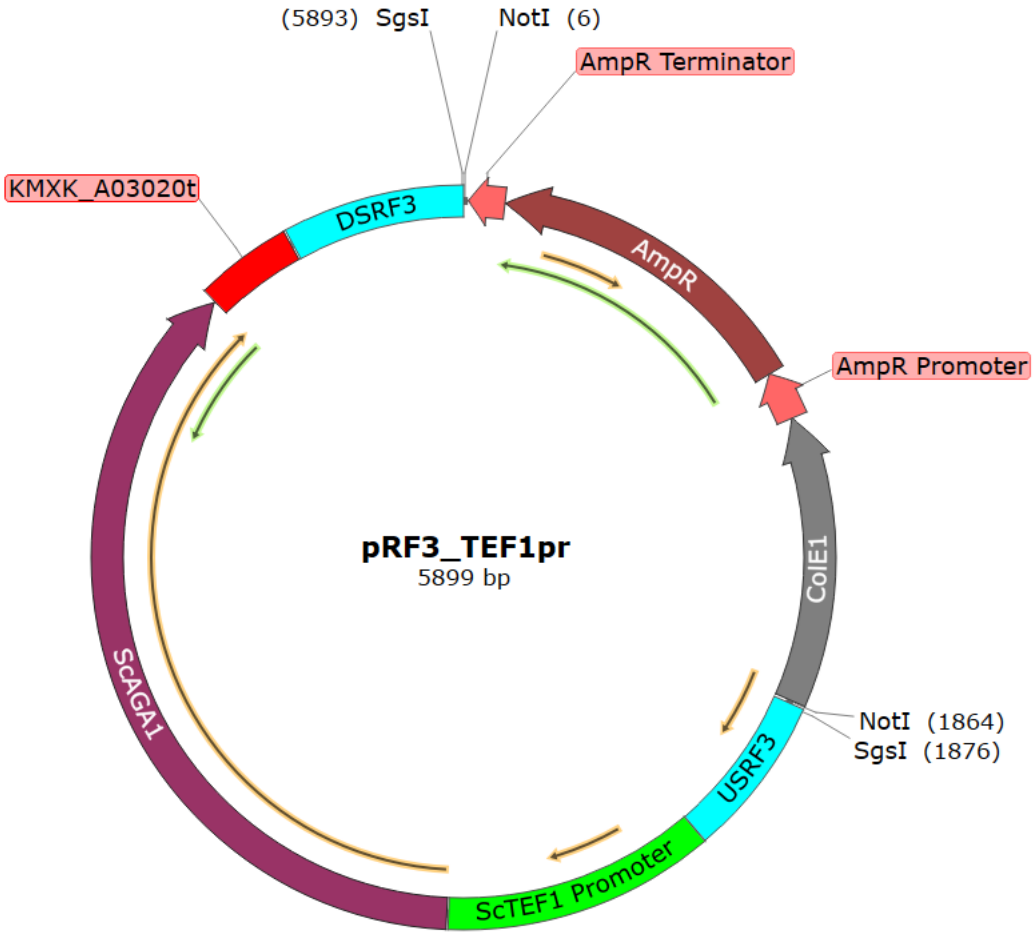


Fig. S2. GG plasmid used to construct YSD platform strain. USRF3 (Yeast Toolkit part 1) sequence corresponds to *KmAGA2* Upstream region, *ScTEF1* promoter (pYTK013, part 2), *ScAGA1* gene (part 3), *KMXK_A03020* terminator (*K. marxianus* toolkit expansion, part 4), DSRF3 (part 5) sequence corresponds to *KmAGA2* Downstream region, Ampicillin resistance and Bacterial origin of replication (pYTK095, part 678).

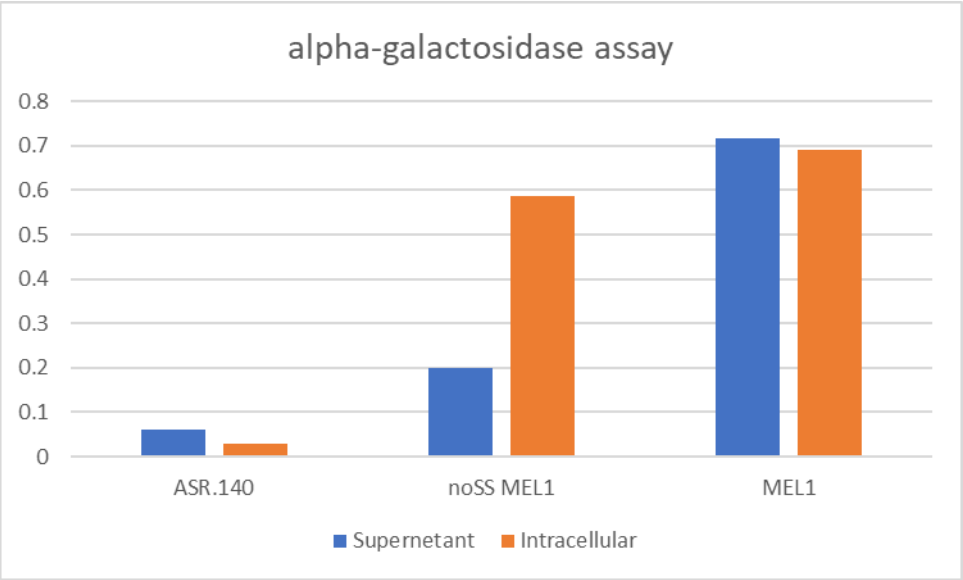


Fig. S3. α-galactosidase activity assay performed on supernatant (blue) and cell pellet (orange) of same yeast culture. ASR.140 (WT), noSS MEL1 (WT expressing *ScMEL1* without SS), MEL1 (WT expressing *ScMEL1* with native SS).

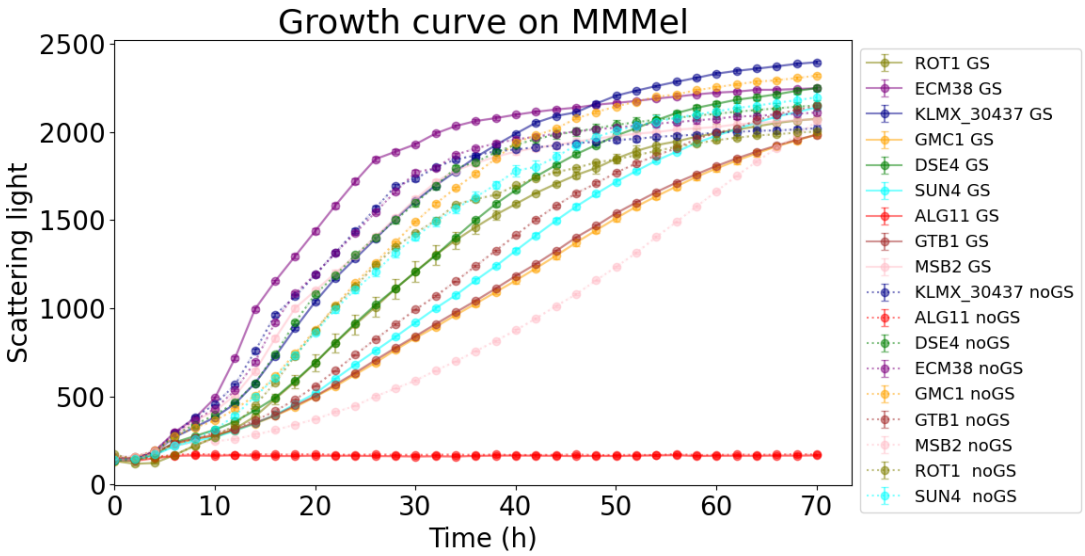


Fig. S4. Growth curve comparison of wild type strain expressing *ScMEL1* fused to the SSs used in this study with either the Gly – Ser linkage introduced by YTK (SS GS, solid lines) or without it (SS noGS, dotted lines). Each colour represents one SS.

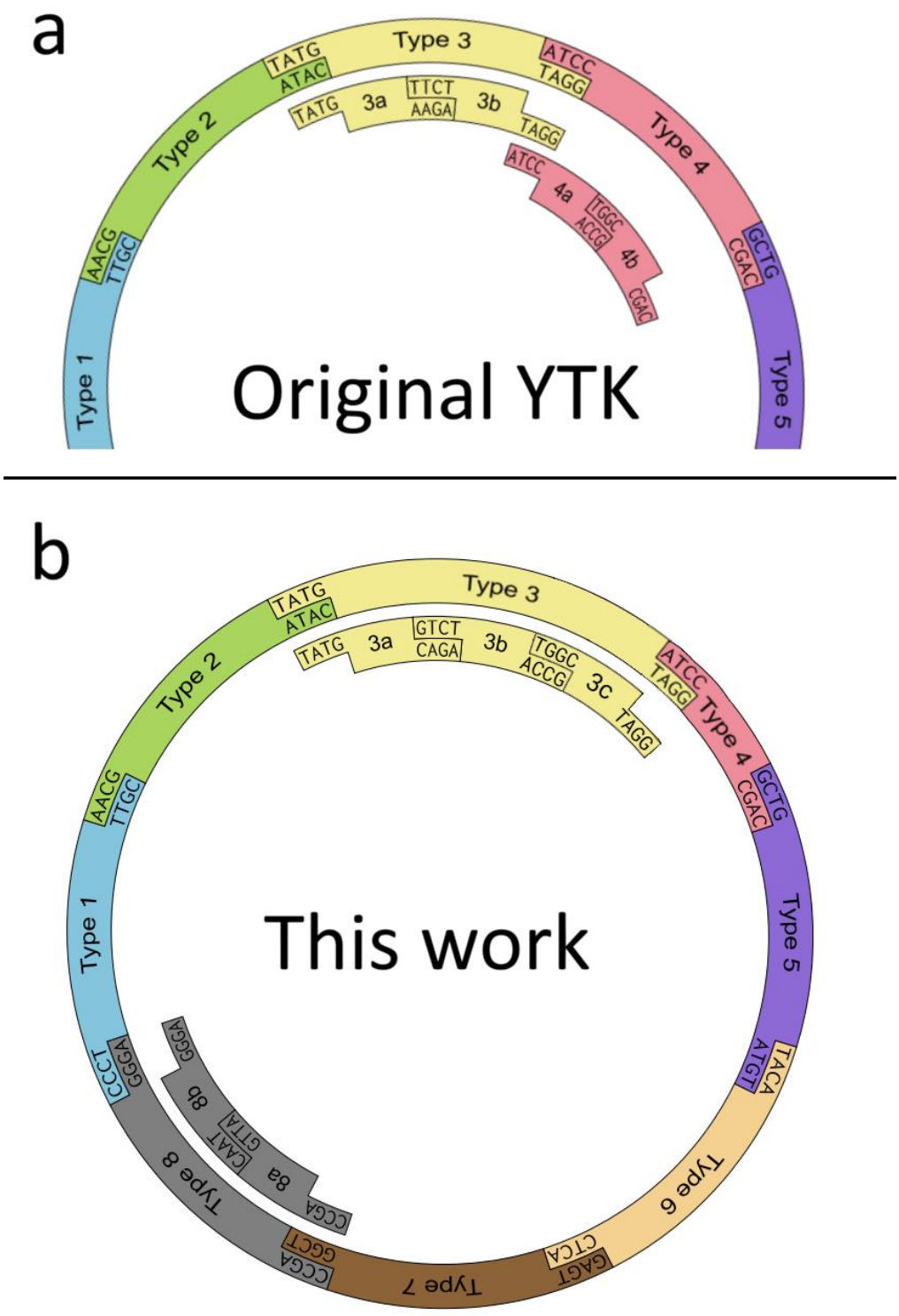


Fig. S5. a) Original YTK overhang configuration, with fusion between type 3b and type 4a parts. **b)** Modified version in this work, where a modified type 3b** part can anneal with the newly introduced type 3c part and the type 3c part can now directly link with type 3 parts.

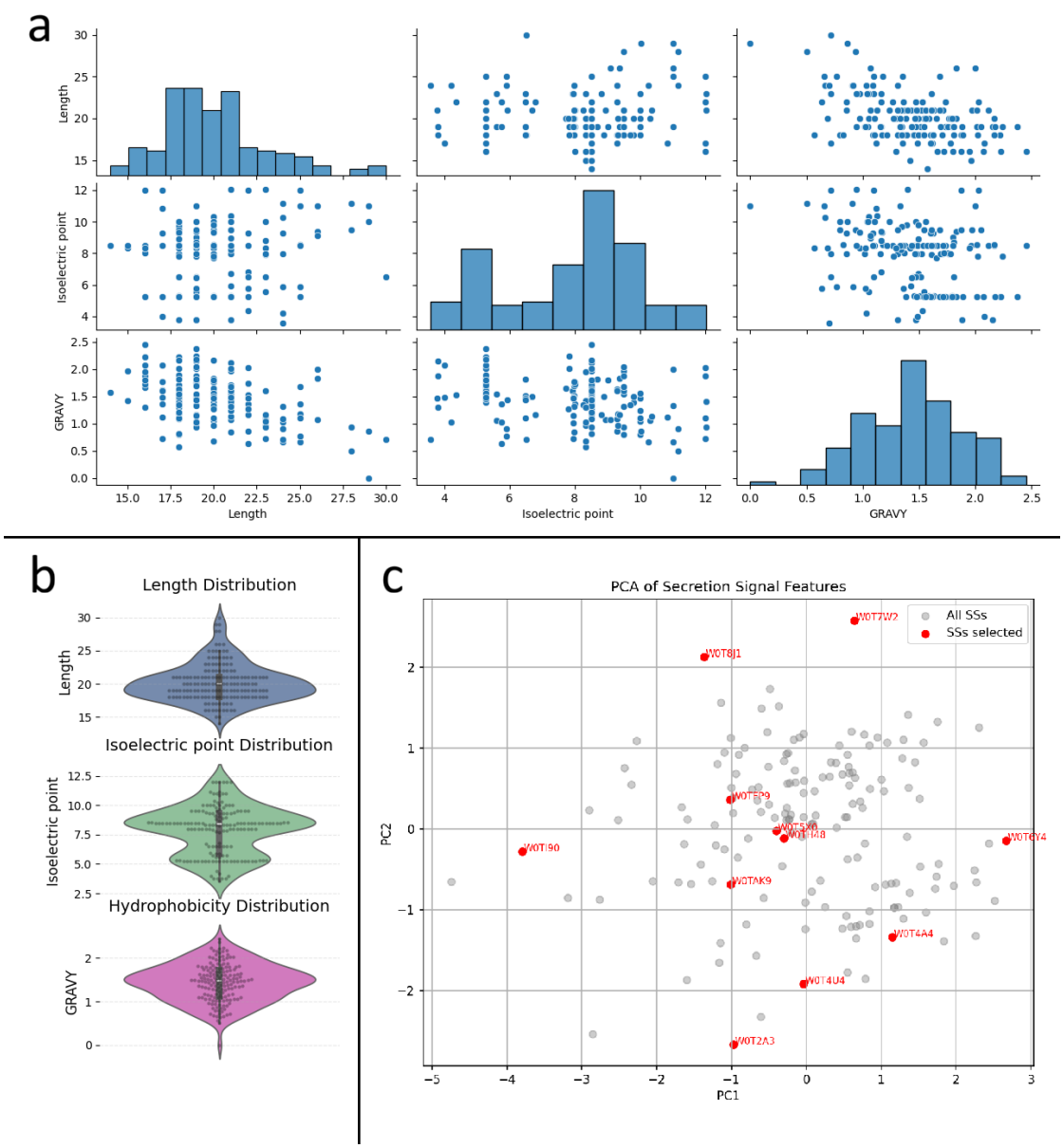


Fig. S6. a) Pairwise relationships between the three features studied in this work: Length, Isoelectric point, GRAVY, of all the SSs of *K. marxianus*. Diagonal shows histograms displaying the frequency distribution of a single feature. Off-diagonal shows scatter plots displaying the correlation between two different variables where each dot represents a SS. **b)** Violin plots showing the distribution of all the SSs for each of the three features. Each dot represents one SS. **c)** PCA displaying all the SSs across the 2-dimensional Principal Component (PC) space. Red dots represent the SSs selected for this study together with their UniProt identifier. Grey dots represent the background population of SSs.

Name	Genotype	Comments
ASR.140 (WT)	ASR.005 I5::CBS-LAC	NBRC 1777 with deletion of lig4 to knockout NHEJ (ASR.005). Resistant to G418 due to insertion of CBS 397 LAC4-LAC12 cluster
Venus	ASR.140 I4::PDC1p-Venus-PDC1t-HphMX	Control strain with no secretion for fluorescence experiments
YSD platform	ASR.140 ΔAGA2::ScAGA1	Platform strain for YSD. AGA2 has been deleted and ScAGA1 under control of ScTEF1 promoter and KMXK_A03020 terminator has been inserted in its place
MEL1	ASR.140 I4::PDC1p-MEL1-PDC1t-HphMX	Control MEL1 strain. Secretes α-galactosidase with native gene untouched
MEL1-GS	ASR.140 I4::PDC1p-MEL1ss-noSSMEL1-PDC1t-HphMX	Control MEL1 strain. Reconstructed by joining MEL1 signal and MEL1 gene without signal. Contains Ser-Gly bond introduced by YTK
noSSMEL1	ASR.140 I4::PDC1p-noSSMEL1-PDC1t-HphMX	Control MEL1 strain. MEL1 has no secretion signal and cannot be secreted, but shows intracellular α-galactosidase activity
αMF noSSMEL1	ASR.140 I4::PDC1p-αMFss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by αMF signal fused to MEL1. Contains Ser-Gly bond introduced by YTK
ROT1 GS	ASR.140 I4::PDC1p-ROT1ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by ROT1 signal fused to MEL1. Contains Ser-Gly bond introduced by YTK
ECM38 GS	ASR.140 I4::PDC1p-ECM38ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by ECM38 signal fused to MEL1. Contains Ser-Gly bond introduced by YTK
KLMX_30437 GS	ASR.140 I4::PDC1p-KLMX_30437ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by KLMX_30437 signal fused to MEL1. Contains Ser-Gly bond introduced by YTK
GMC1 GS	ASR.140 I4::PDC1p-GMC1ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by GMC1 signal fused to MEL1. Contains Ser-Gly bond introduced by YTK
DSE4 GS	ASR.140 I4::PDC1p-DSE4ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by DSE4 signal fused to MEL1. Contains Ser-Gly bond introduced by YTK
SUN4 GS	ASR.140 I4::PDC1p-SUN4ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by SUN4 signal fused to MEL1. Contains Ser-Gly bond introduced by YTK
ALG11 GS	ASR.140 I4::PDC1p-ALG11ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by ALG11 signal fused to MEL1. Contains Ser-Gly bond introduced by YTK
GTB1 GS	ASR.140 I4::PDC1p-GTB1ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by GTB1 signal fused to MEL1. Contains Ser-Gly bond introduced by YTK
MSB2 GS	ASR.140 I4::PDC1p-MSB2ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by MSB2 signal fused to MEL1. Contains Ser-Gly bond introduced by YTK
MEL1 noGS	ASR.140 I4::PDC1p-MEL1ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by MEL1 signal fused to MEL1 without Ser-Gly bond
KLMX_30437 noGS	ASR.140 I4::PDC1p-KLMX_30437ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by KLMX_30437 signal fused to MEL1 without Ser-Gly bond
ALG11 noGS	ASR.140 I4::PDC1p-ALG11ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by ALG11 signal fused to MEL1 without Ser-Gly bond

DSE4 noGS	ASR.140 I4::PDC1p-DSE4ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by DSE4 signal fused to MEL1 without Ser-Gly bond
ECM38 noGS	ASR.140 I4::PDC1p-ECM38ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by ECM38 signal fused to MEL1 without Ser-Gly bond
GMC1 noGS	ASR.140 I4::PDC1p-GMC1ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by GMC1 signal fused to MEL1 without Ser-Gly bond
GTB1 noGS	ASR.140 I4::PDC1p-GTB1ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by GTB1 signal fused to MEL1 without Ser-Gly bond
MSB2 noGS	ASR.140 I4::PDC1p-MSB2ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by MSB2 signal fused to MEL1 without Ser-Gly bond
ROT1 noGS	ASR.140 I4::PDC1p-ROT1ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by ROT1 signal fused to MEL1 without Ser-Gly bond
SUN4 noGS	ASR.140 I4::PDC1p-SUN4ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by SUN4 signal fused to MEL1 without Ser-Gly bond
KLMX_10834 noGS	ASR.140 I4::PDC1p-KLMX_10834ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by KLMX_10834 signal fused to MEL1 without Ser-Gly bond
KRE1 noGS	ASR.140 I4::PDC1p-KRE1ss-noSSMEL1-PDC1t-HphMX	MEL1 strain secreting α-galactosidase activity by KRE1 signal fused to MEL1 without Ser-Gly bond
noSSMEL1	CBM002 I4::PDC1p-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product MEL1 without secretion signal
KLMX_10834 YSD	CBM002 I4::PDC1p-KLMX_10834ss-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product KLMX_10834 secretion signal fused to noSS_MEL1
KLMX_30437 YSD	CBM002 I4::PDC1p-KLMX_30437ss-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product KLMX_30437 secretion signal fused to noSS_MEL1
ALG11 MEL1 YSD	CBM002 I4::PDC1p-ALG11ss-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product ALG11 secretion signal fused to noSS_MEL1
DSE4 MEL1 YSD	CBM002 I4::PDC1p-DSE4ss-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product DSE4 secretion signal fused to noSS_MEL1
ECM38 MEL1 YSD	CBM002 I4::PDC1p-ECM38ss-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product ECM38 secretion signal fused to noSS_MEL1
GMC1 MEL1 YSD	CBM002 I4::PDC1p-GMC1ss-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product GMC1 secretion signal fused to noSS_MEL1
GTB1 MEL1 YSD	CBM002 I4::PDC1p-GTB1ss-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product GTB1 secretion signal fused to noSS_MEL1
KRE1 MEL1 YSD	CBM002 I4::PDC1p-KRE1ss-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product KRE1 secretion signal fused to noSS_MEL1

MEL1 MEL1 YSD	CBM002 I4::PDC1p-MEL1ss-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product MEL1 secretion signal fused to noSS_MEL1
MSB2 MEL1 YSD	CBM002 I4::PDC1p-MSB2ss-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product MSB2 secretion signal fused to noSS_MEL1
ROT1 MEL1 YSD	CBM002 I4::PDC1p-ROT1ss-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product ROT1 secretion signal fused to noSS_MEL1
SUN4 MEL1 YSD	CBM002 I4::PDC1p-SUN4ss-noSSMEL1-ScAGA2-HA-TDH1t-HphMX	Strain expressing YSD fusion product SUN4 secretion signal fused to noSS_MEL1
αMF Venus YSD	CBM002 I4::PDC1p-αMFss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product αMF secretion signal fused to Venus
KLMX_10834 Venus YSD	CBM002 I4::PDC1p-KLMX_10834ss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product KLMX_10834 secretion signal fused to Venus
KLMX_30437 Venus YSD	CBM002 I4::PDC1p-KLMX_30437ss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product KLMX_30437 secretion signal fused to Venus
ALG11 Venus YSD	CBM002 I4::PDC1p-ALG11ss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product ALG11 secretion signal fused to Venus
DSE4 Venus YSD	CBM002 I4::PDC1p-DSE4ss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product DSE4 secretion signal fused to Venus
ECM38 Venus YSD	CBM002 I4::PDC1p-ECM38ss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product ECM38 secretion signal fused to Venus
GMC1 Venus YSD	CBM002 I4::PDC1p-GMC1ss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product GMC1 secretion signal fused to Venus
GTB1 Venus YSD	CBM002 I4::PDC1p-GTB1ss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product GTB1 secretion signal fused to Venus
KRE1 Venus YSD	CBM002 I4::PDC1p-KRE1ss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product KRE1 secretion signal fused to Venus
MEL1 Venus YSD	CBM002 I4::PDC1p-MEL1ss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product MEL1 secretion signal fused to Venus
MSB2 Venus YSD	CBM002 I4::PDC1p-MSB2ss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product MSB2 secretion signal fused to Venus
ROT1 Venus YSD	CBM002 I4::PDC1p-ROT1ss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product ROT1 secretion signal fused to Venus
SUN4 Venus YSD	CBM002 I4::PDC1p-SUN4ss-Venus-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product SUN4 secretion signal fused to Venus
αMF Venus	ASR.140 I4::PDC1p-αMFss-Venus-PDC1t-HphMX	Fluorescent strain expressing αMF secretion signal fused to Venus

KLMX_10834 Venus	ASR.140 I4::PDC1p-KLMX_10834ss-Venus-PDC1t-HphMX	Fluorescent strain expressing KLMX_10834 secretion signal fused to Venus
KLMX_30437 Venus	ASR.140 I4::PDC1p-KLMX_30437ss-Venus-PDC1t-HphMX	Fluorescent strain expressing KLMX_30437 secretion signal fused to Venus
ALG11 Venus	ASR.140 I4::PDC1p-ALG11ss-Venus-PDC1t-HphMX	Fluorescent strain expressing ALG11 secretion signal fused to Venus
DSE4 Venus	ASR.140 I4::PDC1p-DSE4ss-Venus-PDC1t-HphMX	Fluorescent strain expressing DSE4 secretion signal fused to Venus
ECM38 Venus	ASR.140 I4::PDC1p-ECM38ss-Venus-PDC1t-HphMX	Fluorescent strain expressing ECM38 secretion signal fused to Venus
GMC1 Venus	ASR.140 I4::PDC1p-GMC1ss-Venus-PDC1t-HphMX	Fluorescent strain expressing GMC1 secretion signal fused to Venus
GTB1 Venus	ASR.140 I4::PDC1p-GTB1ss-Venus-PDC1t-HphMX	Fluorescent strain expressing GTB1 secretion signal fused to Venus
KRE1 Venus	ASR.140 I4::PDC1p-KRE1ss-Venus-PDC1t-HphMX	Fluorescent strain expressing KRE1 secretion signal fused to Venus
MEL1 Venus	ASR.140 I4::PDC1p-MEL1ss-Venus-PDC1t-HphMX	Fluorescent strain expressing MEL1 secretion signal fused to Venus
MSB2 Venus	ASR.140 I4::PDC1p-MSB2ss-Venus-PDC1t-HphMX	Fluorescent strain expressing MSB2 secretion signal fused to Venus
ROT1 Venus	ASR.140 I4::PDC1p-ROT1ss-Venus-PDC1t-HphMX	Fluorescent strain expressing ROT1 secretion signal fused to Venus
SUN4 Venus	ASR.140 I4::PDC1p-SUN4ss-Venus-PDC1t-HphMX	Fluorescent strain expressing SUN4 secretion signal fused to Venus
αMF CSF2 YSD	CBM002 I4::PDC1p-αMFss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product αMF secretion signal fused to CSF2
KLMX_10834 CSF2 YSD	CBM002 I4::PDC1p-KLMX_10834ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product KLMX_10834 secretion signal fused to CSF2
KLMX_30437 CSF2 YSD	CBM002 I4::PDC1p-KLMX_30437ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product KLMX_30437 secretion signal fused to CSF2
ALG11 CSF2 YSD	CBM002 I4::PDC1p-ALG11ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product ALG11 secretion signal fused to CSF2
CSF2 CSF2 YSD	CBM002 I4::PDC1p-CSF2ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product CSF2 secretion signal fused to CSF2
DSE4 CSF2 YSD	CBM002 I4::PDC1p-DSE4ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product DSE4 secretion signal fused to CSF2

ECM38 CSF2 YSD	CBM002 I4::PDC1p-ECM38ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product ECM38 secretion signal fused to CSF2
GMC1 CSF2 YSD	CBM002 I4::PDC1p-GMC1ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product GMC1 secretion signal fused to CSF2
GTB1 CSF2 YSD	CBM002 I4::PDC1p-GTB1ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product GTB1 secretion signal fused to CSF2
KRE1 CSF2 YSD	CBM002 I4::PDC1p-KRE1ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product KRE1 secretion signal fused to CSF2
MEL1 CSF2 YSD	CBM002 I4::PDC1p-MEL1ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product MEL1 secretion signal fused to CSF2
MSB2 CSF2 YSD	CBM002 I4::PDC1p-MSB2ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product MSB2 secretion signal fused to CSF2
ROT1 CSF2 YSD	CBM002 I4::PDC1p-ROT1ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product ROT1 secretion signal fused to CSF2
SUN4 CSF2 YSD	CBM002 I4::PDC1p-SUN4ss-CSF2-AGA2-HA-PDC1t-HphMX	Strain expressing YSD fusion product SUN4 secretion signal fused to CSF2

Table S1. List of strains used in this work.

Name	Sequence	Comments
STMel1_3a_F	GCATCGTCTCATCGGTCTCATATGTTTGCTTTCTACTTTCTCACCGCATGCATCAGTTTGAAGGGCG	Forward primer to generate secretion tag from native MEL1. No template, use with STMel1_3a_R
STMel1_3a_R	ATGCCGTCTCAGGTCTCAAGAACCCCCAAAAACGCCCTTCAAACTGATGCATGCGGTGAGAAAGTAGAAAGCAAAC	Reverse primer to generate secretion tag from native MEL1. No template, use with STMel1_3a_F
STKLMX_30437_F	GCATCGTCTCATCGGTCTCATATGAGATTTAATAGAATTGCTCCATTTATGTTGGCTCCATTGGCTTCTGCT	Forward primer to generate secretion tag from KLMX_30437 gene. No template, use with STKLMX_30437_R
STKLMX_30437_R	ATGCCGTCTCAGGTCTCAAGAACCAGCAGAAGCCAATGGAGCCAACATAAATGGAGCAATTCTATTAAATCTCAT	Reverse primer to generate secretion tag from KLMX_30437 gene. No template, use with STKLMX_30437_F
STALG11_F	GCATCGTCTCATCGGTCTCATATGATGTCTTTGGCTAAATTTATTACTTTTTTTTTGGGTTTGTTGTTGTTTTTGTTGGTTTCTTTGAGAGTTTTGTCT	Forward primer to generate secretion tag from ALG11 gene. No template, use with STALG11_R
STALG11_R	ATGCCGTCTCAGGTCTCAAGAACCAGACAAAACTCTCAAAGAAACCAACAAAAACAACAACAAACCCAAAAAAAAAGTAATAAATTTAGCCAAAGACATCAT	Reverse primer to generate secretion tag from ALG11 gene. No template, use with STALG11_F
STDSE4_F	GCATCGTCTCATCGGTCTCATATGTTGGCTCAAGTTTGGTTGTTGTCTTTGGGTTATGTTTTGACTACTTTTTTGTCTGCTGTTGGTGTTGCTGCT	Forward primer to generate secretion tag from DSE4 gene. No template, use with STDSE4_R
STDSE4_R	ATGCCGTCTCAGGTCTCAAGAACCAGCAGCAACACCAACAGCAGACAAAAAAGTAGTCAAAACATAACCCAAAGACAACAACCAAACTTGAGCCAACAT	Reverse primer to generate secretion tag from DSE4 gene. No template, use with STDSE4_F
STECM38_F	GCATCGTCTCATCGGTCTCATATGTTGTTTTCTACTCCATATTTGTGGTTGTTGTTGTTGGGTTCTGTTGATGCT	Forward primer to generate secretion tag from ECM38 gene. No template, use with STECM38_R
STECM38_R	ATGCCGTCTCAGGTCTCAAGAACCAGCATCAACAGAACCCAACAACAACAACCACAAATATGGAGTAGAAAACAACAT	Reverse primer to generate secretion tag from ECM38 gene. No template, use with STECM38_F
STGMC1_F	GCATCGTCTCATCGGTCTCATATGGAATTTCCAAATTGGTGGTGTGTTTTTGGTTTGTTGTTGGCTTGGTTGTTGTCTGATACTTGTGAAGGT	Forward primer to generate secretion tag from GMC1 gene. No template, use with STGMC1_R
STGMC1_R	ATGCCGTCTCAGGTCTCAAGAACCACCTTCACAAGTATCAGACAACAACCAAGCCAACAACAAACCAAAAACACACCACCAATTTGGAAATTCCAT	Reverse primer to generate secretion tag from GMC1 gene. No template, use with STGMC1_F
STGTB1_F	GCATCGTCTCATCGGTCTCATATGTTGTTTCCAATTGCTTTTTTGGCTTGTGTTGTTTGTGCTTCTGGT	Forward primer to generate secretion tag from GTB1 gene. No template, use with STGTB1_R
STGTB1_R	ATGCCGTCTCAGGTCTCAAGAACCACCAGAAGCACAAACAACACAAGCCAAAAAAGCAATTGGAAACAACAT	Reverse primer to generate secretion tag from GTB1 gene. No template, use with STGTB1_F
STMSB2_F	GCATCGTCTCATCGGTCTCATATGCAATTTTCTTCTATTAAAGTTGGTTTGGCTTTGTTGGTTTTGCAATCTTGTGCTGCTTCTGTTTCTGCT	Forward primer to generate secretion tag from MSB2 gene. No template, use with STMSB2_R
STMSB2_R	ATGCCGTCTCAGGTCTCAAGAACCAGCAGAAACAGAAGCAGCACAAGATTGCAAAACCAACAAAGCCAAACCAACTTTAATAGAAGAAAATTGCA	Reverse primer to generate secretion tag from MSB2 gene. No template, use with STMSB2_F
STROT1_F	GCATCGTCTCATCGGTCTCATATGAGATTTTTGATTTTGTTGTCTATTTTTGCTGTTGCTATTAGAGGT	Forward primer to generate secretion tag from ROT1 gene. No template, use with STROT1_R

STROT1_R	ATGCCGTCTCAGGTCTCAAGAACCACCTCTAATAGCAACAGCAAAAATAGACAACAAAATCAAAAATCTCAT	Reverse primer to generate secretion tag from ROT1 gene. No template, use with STROT1_F
STSUN4_F	GCATCGTCTCATCGGTCTCATATGAAATTTGCTACTGCTTTGGCTTCTGCTTCTTTTATTGGTGCTACTTTGGTTTCTGCTTTGCCACATCCACATGCTAAAAGA	Forward primer to generate secretion tag from SUN4 gene. No template, use with STSUN4_R
STSUN4_R	ATGCCGTCTCAGGTCTCAAGAACCTCTTTTAGCATGTGGATGTGGCAAAGCAGAAACCAAAGTAGCACCAATAAAAGAAGCAGAAGCCAAAGCAGTAGCAAATTTCAT	Reverse primer to generate secretion tag from SUN4 gene. No template, use with STSUN4_F
AGA2_Diag_F	cgaaaccggccatctaga	Forward primer to amplify KmAGA2. For CRISPR diagnosis. To use with AGA2_Diag_R
AGA2_Diag_R	ctagactgtctgggtaggga	Reverse primer to amplify KmAGA2. For CRISPR diagnosis. To use with AGA2_Diag_R
AGA2_TF	AAACgtggaacaattccaagccca	CRISPR/Cas9 target primer F
AGA2_TR	CGTCtgggcttggaattgttccac	CRISPR/Cas9 target primer R
ScAGA1_Diag_F	gccacgagtacatcaagcat	Forward primer annealing in ScAGA1 to amplify insertion in KmAGA2. For CRISPR diagnosis. To use with AGA2_Diag_R
ScAGA1_Diag_R	gtcaatggacaccaagtcgt	Reverse primer annealing in ScAGA1 to amplify insertion in KmAGA2. For CRISPR diagnosis. To use with AGA2_Diag_F
AGA2_USRF3_SgsI_F	GCATCGTCTCATCGGTCTCACCCTGGCGCGCCgaccaacggtgctaccgcta	Forward primer to amplify region upstream AGA2 gene as a YTK type 1 part. For disruption and integration in KmAGA2. Contains SgsI site for linearisation of repair fragment. To use with AGA2_USRF3_F
AGA2_USRF3_R	ATGCCGTCTCAGGTCTCACGTTgtatataggattctggtaatgca	Reverse primer to amplify region upstream AGA2 gene as a type 1 part for disruption and integration in AGA2. To use with AGA2_DSRF3_SgsI_F
AGA2_DSRF3_F	GCATCGTCTCATCGGTCTCAGCTGtgatgctgtatttggagtca	Forward primer to amplify region downstream AGA2 gene as a type 1 part for disruption and integration in AGA2. To use with AGA2_DSRF3_SgsI_F
AGA2_DSRF3_SgsI_R	ATGCCGTCTCAGGTCTCATGTAGGCGCGcctgctatttcctccagcat	Reverse primer to amplify region downstream AGA2 gene as a type 1 part for disruption and integration in AGA2. Contains SgsI site for linearisation of repair fragment. To use with AGA2_DSRF3_R
Mel1_F	GCATCGTCTCATCGGTCTCATATGTTTGCTTTCTACTTTCTCACCGC	Forward primer for amplification of MEL1 gene as a type 3 part for GG assembly
Mel1_R	ATGCCGTCTCAGGTCTCAGGATTCAAGAAGAGGGccgcAAC	Reverse primer for amplification of MEL1 gene as a type 3/3b part for GG assembly

ScMEL1_noST_F	GCATCGTCTCATCGGTCTCATATGGTGTCTCCGAGTTACAATGGCC	Forward primer for generation of MEL1 gene type 3 part without native Secretion signal
ScMEL1_noST_3b_F	GCATCGTCTCATCGGTCTCATTCTGTGTCTCCGAGTTACAATGGCC	Forward primer for generation of MEL1 gene type 3b part without native Secretion signal
MEL1_noSTOP_level1_R	ATGCCGTCTCAGGTCTCAGGATAGAAGAGGGccgcAACCT	Reverse primer for generation of MEL1 gene type 3b part without native Secretion signal and without STOP codon
ScMEL1_noST_3b_NewOH_F	GCATCGTCTCATCGGTCTCAGTCTCCGAGTTACAATGGCCTTGGCCTT	Forward primer for generation of MEL1 gene type 3b part without native Secretion signal. When fused to a 3a part does not produce Gly-Ser bond between 3a and 3b parts. To use with D9.
ScMEL1_noST_3a_NewOH_F	GCATCGTCTCATCGGTCTCATATGGTGTCTCCGAGTTACAATGGCC	Forward primer to amplify ScMEL1 gene without secretion tag as a type 3 part
MEL1_AGA2sys_F	GCATCGTCTCATCGGTCTCAGTCTCCGAGTTACAATGGCCTT	Forward primer to amplify ScMEL1 without Secretion signal and without STOP codon. New overhangs, not fully compatible with YTK
MEL1_AGA2sys_R	ATGCCGTCTCAGGTCTCAGCCAAGAAGAGGGCCGCAACCT	Reverse primer to amplify ScMEL1 without Secretion signal and without STOP codon. New overhangs, not fully compatible with YTK
AGA2_HA_F	GCATCGTCTCATCGGTCTCATGGCGTGGTAGTGGGTCCGGTTCT	Forward primer to amplify construct AGA2-HA as a fragment for level I GG assembly. To use with AGA2_HA_R
AGA2_HA_R	ATGCCGTCTCAGGTCTCAGGATGTTAAGCGTAGTCTGGAACGTCGT	Reverse primer to amplify construct AGA2-HA as a fragment for level I GG assembly. To use with AGA2_HA_F
Venus_agasys_F	GCATCGTCTCATCGGTCTCAGTCTtctaaaggtgaagaattattcactggt	Forward primer to amplify Venus (YTK) to adapt for YSD. To use with Venus_agasys_R
Venus_agasys_R	ATGCCGTCTCAGGTCTCAGCCAtcctttgtacaattcatccatacca	Reverse primer to amplify Venus (YTK) to adapt for YSD. To use with Venus_agasys_F
ST_Hs_CSF2_F	GCATCGTCTCATCGGTCTCATATGTGGCTACAGTCGCTACTACTACTAGGTACCGTT	Forward primer to generate secretion tag from KRE1 gen without gly-ser bond typical from GG. No template, use with ST_Hs_CSF2_R
ST_Hs_CSF2_R	ATGCCGTCTCAGGTCTCAAGACGATATCGAGCATGCAACGGTACCTAGTAGTAGTA	Forward primer to generate secretion tag from KRE1 gen without gly-ser bond typical from GG. No template, use with ST_Hs_CSF2_F

Table S2. Primers used in this work.

Name	References & Comments	Features
pI4-MTU-DO-HPH	Integrative preassembled vector for Golden Gate assembly. Targets I4 (I3 in the Frontiers paper) locus (ChrIV: 388650..390345) with a hygromycin resistance marker.	KanR, ColE
pYTK001	Level I storage plasmid empty. For cloning other level I parts.	CamR, ColE
pP2	Level I storage plasmid containing KmPDC1 promoter from KmK toolkit.	CamR, ColE
pYTK033	Level I storage plasmid containing Venus from Yeast ToolKit.	CamR, ColE
pVenus_agasys	Level I storage plasmid containing Venus (3b part) amplified from YTK045 with different overhangs to adapt to YSD. Compatible with pAGA2_HA.	CamR, ColE
pYTK045	Level I storage plasmid containing Venus (3b part) from YTK Kit.	CamR, ColE
pMEL1	Level I storage plasmid containing alpha-galactosidase from S.cerevisiae MEL1 gene.	CamR, ColE
pMEL1_noSS_3	Level I storage plasmid containing S.cerevisiae MEL1 gene. Native secretion signal removed. Type 3 part.	CamR, ColE
pMEL1_noSS_3b	Level I storage plasmid containing S.cerevisiae MEL1 gene. Native secretion signal removed. Type 3b part.	CamR, ColE
pnoSSMEL1noOH	Level I storage plasmid containing S.cerevisiae MEL1 gene. Native secretion signal removed. Modified overhangs to avoid Gly-Ser bond. Type 3b part.	CamR, ColE
pScMEL1_noSTOP_t3	Level I storage plasmid containing noSSMEL1 adapted for YSD. Compatible with YTK promoters but not terminators. To be used with pAGA2-HA.	CamR, ColE
pScMEL1_agasys	Level I storage plasmid containing noSSMEL1 adapted for YSD. Not compatible with YTK. To be used with pAGA2-HA.	CamR, ColE
pScAGA1	Level I storage plasmid containing S. cerevisiae AGA1 gene.	CamR, ColE
pnoSSCSF2	Level I storage plasmid containing Human CSF2 as a 3b part. Not compatible with YTK. Compatible to fuse with pAGA2_HA	CamR, ColE
pAGA2-HA	Level I storage plasmid containing AGA2 fused to HA antigen adapted for YSD fusions. Not compatible with YTK genes. Compatible with YTK terminators.	CamR, ColE
pPTK005	Level I storage plasmid containing Sc mating factor with YTK promoter overhangs for protein secretion. From Pichia ToolKit.	CamR, ColE
pSSKLMX_30437	Level I storage plasmid containing KLMX_30437 secretion signal.	CamR, ColE
pSSALG11	Level I storage plasmid containing ALG11 secretion signal.	CamR, ColE
pSSDSE4	Level I storage plasmid containing DSE4 secretion signal.	CamR, ColE
pSSECM38	Level I storage plasmid containing ECM38 secretion signal.	CamR, ColE
pSSGMC1	Level I storage plasmid containing GMC1 secretion signal.	CamR, ColE
pSSGTB1	Level I storage plasmid containing GTB1 secretion signal.	CamR, ColE

pSSMEL1	Level I storage plasmid containing MEL1 secretion signal.	CamR, ColE
pSSMSB2	Level I storage plasmid containing MSB2 secretion signal.	CamR, ColE
pSSROT1	Level I storage plasmid containing ROT1 secretion signal.	CamR, ColE
pSSSUN4	Level I storage plasmid containing SUN4 secretion signal.	CamR, ColE
pSS10834noOH	Level I storage plasmid containing KLMX_10834 secretion signal that generates no Ser-Gly GG bond.	CamR, ColE
pSS30437noOH	Level I storage plasmid containing KLMX_30437 secretion signal that generates no Ser-Gly GG bond.	CamR, ColE
pSSALG11noOH	Level I storage plasmid containing ALG11 secretion signal that generates no Ser-Gly GG bond.	CamR, ColE
pSS_CSF2	Level I storage plasmid containing Human CSF2 secretion signal. Generates Valine-Serine bond (from MEL1 overhangs).	CamR, ColE
pSSDSE4noOH	Level I storage plasmid containing DSE4 secretion signal that generates no Ser-Gly GG bond.	CamR, ColE
pSSECM38noOH	Level I storage plasmid containing ECM38 secretion signal that generates no Ser-Gly GG bond.	CamR, ColE
pSSGMC1noOH	Level I storage plasmid containing GMC1 secretion signal that generates no Ser-Gly GG bond.	CamR, ColE
pSSGTB1noOH	Level I storage plasmid containing GTB1 secretion signal that generates no Ser-Gly GG bond.	CamR, ColE
pSSKRE1noOH	Level I storage plasmid containing KRE1 secretion signal that generates no Ser-Gly GG bond.	CamR, ColE
pSSMel1noOH	Level I storage plasmid containing secretion signal from MEL1 gene, modified to not form typical Gly-Ser bond (no OverHang) from GG MoClo toolkit. Type 3a part.	CamR, ColE
pSSMSB2noOH	Level I storage plasmid containing MSB2 secretion signal that generates no Ser-Gly GG bond.	CamR, ColE
pSSROT1noOH	Level I storage plasmid containing ROT1 secretion signal that generates no Ser-Gly GG bond.	CamR, ColE
pSSSUN4noOH	Level I storage plasmid containing SUN4 secretion signal that generates no Ser-Gly GG bond.	CamR, ColE
pYTK066	Level I storage plasmid containing YTK066, ScTDH1 terminator as a 4b type part, from Yeast Toolkit.	CamR, ColE
pYTK056	Level I storage plasmid containing ScTDH1 terminator from YTK kit.	CamR, ColE
pKmK.T004	Level I storage plasmid containing KmPDC1 terminator from KmK toolkit.	CamR, ColE
pUCC001	Empty CRISPR-Cas9 plasmid for gRNA cloning. HphMX for yeast selection.	AmpR, ColE
pUCC-AGA2T2	CRISPR plasmid for deletion of KmAGA2.	AmpR, ColE
pUSRF3-AGA2-SgsI	Level I storage plasmid containing the sequence upstream of KmAGA2. SgsI site for plasmid linearisation.	CamR, ColE
pDSRF3-AGA2-SgsI	Level I storage plasmid containing the sequence downstream of KmAGA2. SgsI site for plasmid linearisation.	CamR, ColE
pYTK002	Level I storage plasmid containing ConLS part from YTK toolkit.	CamR, ColE
pScAGA1	Level I storage plasmid containing S. cerevisiae AGA1 gene.	CamR, ColE

pYTK095	Level II storage plasmid containing Ampicillin resistance - ColE1 origin of replication from YTK Kit.	AmpR, ColE
pKmK.T003	Level I storage plasmid containing KMXK_A03020 terminator from KmK toolkit.	CamR, ColE
pRF3_TEF1pr	Level II storage plasmid containing a Repair Fragment for CRISPR/Cas9 targeting KmAGA2. The RF contains ScAGA1 under the control of ScTEF1 promoter and KMXK_A03020. terminator	AmpR, ColE
pI4-PDC1p-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - Venus - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-αMF-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - αMF secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KLMX_10834-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KLMX_10834 secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KLMX_30437-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KLMX_30437 secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ALG11-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - ALG11 secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-DSE4-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - DSE4 secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ECM38-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter -ECM38 secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-GMC1-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - GMC1 secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-GTB1-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - GTB1 secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KRE1-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KRE1 secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-MEL1-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - MEL1 secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-MSB2-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - MSB2 secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ROT1-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - ROT1 secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-SUN4-Venus-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - SUN4 secretion signal - Venus from YTK kit - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-αMF-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - αMF secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE

pI4-PDC1p-KLMX_10834-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KLMX_10834 secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KLMX_30437-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KLMX_30437 secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ALG11-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - ALG11 secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-DSE4-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - DSE4 secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ECM38-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - ECM38 secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-GMC1-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - GMC1 secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-GTB1-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - GTB1 secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KRE1-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KRE1 secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-MEL1-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - MEL1 secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-MSB2-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - MSB2 secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ROT1-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - ROT1 secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-SUN4-Venus-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - SUN4 secretion signal - Venus from YTK kit - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-MEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - MEL1 - PDC1 terminator - HphMXMX	KanR, ColE
pI4-PDC1p-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE

pI4-PDC1p-αMF-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - αMF secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KLMX_10834-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KLMX_10834 secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KLMX_30437-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KLMX_30437 secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ALG11-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - ALG11 secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-DSE4-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - DSE4 secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ECM38-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter -ECM38 secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-GMC1-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - GMC1 secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-GTB1-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - GTB1 secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KRE1-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KRE1 secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-MEL1-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - MEL1 secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-MSB2-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - MSB2 secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ROT1-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - ROT1 secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-SUN4-noSSMEL1-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - SUN4 secretion signal - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - MEL1 without secretion signal - PDC1 terminator - HphMX	KanR, ColE

pI4-PDC1p-αMF-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - αMF secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KLMX_10834-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KLMX_10834 secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KLMX_30437-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KLMX_30437 secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ALG11-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - ALG11 secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-DSE4-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - DSE4 secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ECM38-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - ECM38 secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-GMC1-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - GMC1 secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-GTB1-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - GTB1 secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KRE1-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KRE1 secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-MEL1-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - MEL1 secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-MSB2-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - MSB2 secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ROT1-noSSMEL1-	Integrative plasmid targeting I4 containing PDC1 promoter - ROT1 secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE

AGA2-HA-PDC1t-hph		
pI4-PDC1p-SUN4-noSSMEL1-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - SUN4 secretion signal - MEL1 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-αMF-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - αMF secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KLMX_10834-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KLMX_10834 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KLMX_30437-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KLMX_30437 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ALG11-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - ALG11 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-CSF2-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - DSE4 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-DSE4-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - DSE4 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ECM38-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - ECM38 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-GMC1-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - GMC1 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-GTB1-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - GTB1 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-KRE1-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - KRE1 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-MEL1-noSSCSF2-	Integrative plasmid targeting I4 containing PDC1 promoter - MEL1 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE

AGA2-HA-PDC1t-hph		
pI4-PDC1p-MSB2-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - MSB2 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-ROT1-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - ROT1 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE
pI4-PDC1p-SUN4-noSSCSF2-AGA2-HA-PDC1t-hph	Integrative plasmid targeting I4 containing PDC1 promoter - SUN4 secretion signal - CSF2 without secretion signal - AGA2-HA fusion product for YSD - PDC1 terminator - HphMX	KanR, ColE

Table S3. All the plasmids used in this study.

Signal Sequence	Protein Code	Length	Isoelectric point	GRAVY
MRFTSLIASASILGAAVSTPIHGHKDNKR	W0TK22	29	11.00	0.000
MKFATALASASFIGATLVSALPHPHAKR	W0TI90	28	11.17	0.504
MQYKNGITAAAILAASVQ	W0TCN1	18	8.34	0.572
MQRWLVFEYLVWFWAFCGSSWA	W0TJB2	22	5.75	0.636
MRLLNVANKVFGTLAIIQGITSANP	W0TDC4	25	11.00	0.656
MRLSDVCVGTLLASVVSGGSISRR	W0T5V4	24	10.26	0.667
MKFTNYLLQAAALLSSQAFA	W0T4X9	20	8.34	0.685
MEFPNWWCVFGLLLAWLLSDTCEG	W0T2A3	24	3.57	0.704
MKGSDMEVQSWPLIRLFVMLIWGHSLLVSA	W0TFS2	30	6.51	0.713
MLNPKVAYMVWMTCLGLTLPSQA	A0A068FCL2	23	7.95	0.717
MRFNRIAPFMLAPLASA	W0T8J1	17	12.00	0.718
MLFSNTLLIAAASALLAEASPLEKR	W0T504	25	5.90	0.768
MKWFNWFAAVLSVRCVSG	W0T9Y1	18	9.50	0.794
MQIKKTFAASALAGSAIA	W0TBA3	18	10.00	0.794
MKFSTVANTALLATFAHA	W0T898	18	8.52	0.822
MRGSGVLQALAVASCASYASA	W0T5M4	21	7.98	0.852
MSWLKISSYVVRTVCLVHITHA	W0T7S9	22	9.30	0.855
MGFGSKRKLKLLLGALFATCLISVFTTYG	W0TAP8	29	10.03	0.866
MRIFTASALKLLPLLATLLHKVQS	W0T597	24	11.17	0.908
MLVRPNECLRWLLWLLTFIVLSEG	W0T8L3	24	5.90	0.913
MTSGHLSRGPLRVILFIFLTFYSICANA	W0TJP1	28	9.50	0.932
MRFKSVHRVLTLLLVMFLAKSTA	W0T7Q1	23	12.02	0.939
MKNVFFPLTTLFLLQVANG	W0TDT4	19	8.50	0.942
MRYSGLTLGALGALALKIAGVEA	W0T8S0	23	8.34	0.970
MKCWKLMLTTLTAFLTVASG	W0TAH6	20	9.31	0.970
MRYSCIASAIALSSSVAYA	W0TC08	19	7.95	1.016
MKFETTLLSVTALSSLAAA	W0TG70	19	5.75	1.021
MKLGGALLSAAFYGQLALS	W0TAZ7	19	8.34	1.021
MQLSGKTLLSLASLFIALLDGADA	W0TFQ2	24	4.21	1.029
MQFKKSAVVSFLSLLGSLSTALA	W0T8S4	23	10.00	1.039
MKKIILFLVPAFHALSTNA	W0TI98	19	10.00	1.047
MSKHLFKPVSVLLFVFKALS	W0TE11	20	10.30	1.050
MTDSYLNIRSLLSLLALLSYALA	W0T4H4	23	5.59	1.052
MKQILKGLMKSFYLSTLLCLVAVVEG	W0T6N5	26	9.10	1.065
MRSDILLRIFFLLLQCSLGKC	W0T9Y6	21	8.94	1.067
MQIKSAAQLLALATAVSA	W0TFN7	18	8.50	1.089
MLSSMKRLLCTVTILLWHISVTYA	W0T6V0	24	9.30	1.092
MTTFLVAFSSVALSSALPCPSHA	W0THP6	23	6.49	1.096
MRAIRSSNLLIIAWLFALKLSIAQA	W0T2S9	25	12.01	1.100
MRYSTLVSATLLAAASSVAA	W0TI51	20	8.50	1.105
MKFSKILLFSSSLTAAFA	W0TDZ2	18	10.00	1.106
MQRLLILLFGLMYHIRA	W0TCB3	17	10.83	1.118
MRQFCSQVVVIILFLLHATRG	W0TAU5	21	10.35	1.129
MKFSSALIYSATVCQVVFS	W0T670	19	7.95	1.153
MSHHILYFLCSVLSLFGNAFVH	W0TI55	22	6.80	1.159
MLKYRCITTYITLLWLLATVAYC	W0T9A2	23	8.80	1.161
MKLFYIFAIVLWYVRCSRC	W0T6A1	19	9.38	1.163
MKLSKTFLIGSFGIPVFC	W0T887	18	9.31	1.172

MRTGFVSMWSGFALLALHLSVFVYG	W0TAF6	25	8.52	1.176
MKFNSVILTSFLTASALA	W0T6G4	18	8.50	1.211
MLFGKTIKLGLATLMAVNAAA	W0T4B9	21	10.00	1.238
MNSKVLQQILIVFSIIVSSVNA	W0T4X3	22	8.50	1.264
MINWSYYALFCLAIVPFAKC	W0TBJ8	20	7.80	1.270
MKFAYSLLLPLAGVSA	P28999	16	8.34	1.294
MKISILSTLLLLGTFDKVIS	W0TH08	20	8.35	1.295
MLFSTPYLWLLLLGSVDA	W0T4A4	18	3.80	1.306
MQFSSIKVGLALLVLQSCAASVSA	W0TAK9	24	7.98	1.308
MKGFSASILLLCFMLLRQAFG	W0TC74	21	9.50	1.310
MKFSTTFALISAAMVSVASA	W0T8Y7	20	8.50	1.330
MKFSIITTIAVTSLSVVRA	W0TDW8	19	11.00	1.332
MKSYYSLIGTILIGIINFAFA	W0T5X0	21	8.25	1.338
MRLGIWSYLILSISSLVSA	W0THH8	19	8.50	1.342
MLFKYYWVSLIFLRLISA	W0TBT4	18	9.70	1.344
MKLSSATLLTLAATAIA	W0T812	17	8.50	1.353
MRLSAVFVSAIALLSTVIAAPITEK	W0TDC1	25	8.50	1.356
MRFSTALGAAISLAVLSTEVVA	W0TEY2	22	5.75	1.364
MRNNVLVSLVALLAACAWVQA	W0TH48	21	8.00	1.367
MQFKNIVAIAAVATVASA	W0T269	18	8.50	1.372
MWFFYLTQLTLMCIIPLVNG	W0TJK5	20	5.27	1.385
MLKCVFKTIALLSVLTKA	W0TCU6	18	9.79	1.389
MKKFGRLLRLLIVVLLAVSSA	W0TAC4	21	12.02	1.400
MKYLTLLTILAVVQG	W0T6Q2	15	8.34	1.420
MLHWHLLLFAIFANTILCVES	W0TE23	21	5.94	1.429
MNSYLHLLLCFGNFLFCLA	W0T9J4	19	6.48	1.432
MKFISTVFSLSAIVSTTVA	W0TD77	19	8.50	1.437
MRLDWWLHFLVLLVASVA	W0T3W8	18	6.50	1.450
MQFSTVATFAAVAAVASA	W0TAY1	18	5.28	1.456
MVSFSKLGSLALLSGAALVSA	W0T6F2	21	8.50	1.457
MKAGLLVSSTLSLFCAVAYA	W0TA42	20	7.95	1.465
MNILKLFITTLLIGKLSFA	W0T892	19	10.00	1.468
MLSSWVSFVLLLLFEFEFTAG	W0TF90	21	3.79	1.471
MQPIGLLSLLFSIVQLMCTLHA	W0TD53	22	6.49	1.473
MKFQLFCLFAYLWQAVFVAA	W0T7X2	20	7.95	1.475
MAYWTAFTTFCIFILTGVVTG	W0T784	21	5.27	1.481
MVKIQLVLLSGLVQLAQVASA	W0TIC8	21	8.50	1.486
MLFNKLAAVSLASLFSSVVTA	W0TFU2	21	8.50	1.486
MLLWLVWFETLWWLALG	W0TFW0	17	4.00	1.488
MKCFIKGLAILSIALKAVSA	W0T946	20	9.79	1.490
MKFSKVCAVLASALCVSA	W0TCC7	18	8.89	1.494
MVMLIHLVYILLLHLTRFADS	W0TIX3	21	6.69	1.495
MTTLLLLSQLGLLAFANA	W0TAJ4	18	5.28	1.506
MQSLLLHAFLLIFLPLSNA	W0TB41	19	6.49	1.516
MKYHTLSSVVCLLISLVVVQA	W0T4N6	21	7.98	1.519
MVRLFDILGALLVAELMGLANA	W0TIA9	22	4.37	1.532
MKISAILSSFAAISAVGA	W0TB02	18	8.50	1.539
MRFLLFFIQSWWWLLVFVTA	W0TG14	20	9.50	1.540
MKLSTLFISLVLALGSVDA	W0TBQ1	19	5.59	1.542
MINQIVLSFIAISGVANA	W0TJP5	18	5.27	1.544

MKFSAAFTVALLSLAKFIAA	W0TFJ8	20	10.00	1.555
MIFQVLLQFFVLLQLPLNAYA	W0T3U4	21	5.27	1.557
MKFGLILLASSAMA	W0TCY1	14	8.50	1.571
MVSSSSFVVALLAWVSLAQA	W0TEI6	20	5.28	1.585
MILSVQWCLLAFVTLVQAVNS	W0T6S4	21	5.27	1.600
MWARWLLVYLASIAITVA	W0TEV2	18	8.50	1.600
MKYILCLLLLFIGRNLA	W0T4H1	17	9.30	1.600
MVYMLKQIISLCVLIYLTGFTLG	W0T9Q3	23	7.93	1.609
MIFNAQFQLLLLALLVKPLLG	W0TCT1	21	8.50	1.610
MIFPLRYCSFLLAFFLRLITA	W0T6N1	21	9.50	1.610
MNLCLFFWFLLASFANA	W0TC68	17	5.27	1.618
MKFTTVAAVALAASTVVA	W0T8P6	18	8.50	1.628
MLKLYICVLIALGRVASG	W0TBN8	18	9.30	1.628
MVTLSLLQTLAFVSFVSVTSA	W0TGA4	21	5.28	1.633
MRVSYLCLFVWIQFICPALC	W0TDN8	20	7.73	1.650
MRLLLLILYLVSGSSA	W0T4V1	16	8.50	1.663
MLAQVWLLSLGYVLTTFLSAVGVAA	W0T4U4	25	5.27	1.680
MFGVNRAFITLFFMLFAVVNA	W0TG08	21	9.50	1.681
MRLLYFVLLLNCGFILVQG	W0T759	19	7.98	1.689
MSIPSIIKLLLFCALANA	W0TJB7	18	7.98	1.706
MILLSNIIFWILYAGVFPSFG	W0TFJ5	21	5.27	1.714
MKSVLGVPLSLLLLLLPLASG	W0T330	21	8.50	1.714
MRLFCLLTWFAGVVFG	W0TIS9	16	8.00	1.719
MYFFTLAALLLFSSRLIAA	W0THM9	19	8.50	1.732
MLTGFIFVCFCILTWLTNTCLG	W0TAR8	22	5.27	1.736
MTIPSFWILVAVLFRAVWC	W0T6N3	19	8.00	1.758
MVLTLLRYLCFWAFFGVSLA	W0TCK1	20	7.98	1.760
MKLVSFAFICWLLSLSIVNA	W0T5N2	20	7.98	1.775
MVYSLGALLLLSSVVNC	W0TBU9	17	5.27	1.776
MVLPTALLAALAATPFVGA	W0T2Y2	19	5.28	1.795
MQYIFALILLLCQLVNPCLA	W0TFP6	20	5.27	1.795
MAGFLTVILGLLWFIQCSFG	W0TBS9	20	5.27	1.795
MKLFVVPVLFASSAFA	W0T663	16	8.50	1.800
MRMLLFAAWLALCVTRVVC	W0THW9	19	9.01	1.800
MLRLPAAICCYLLLCLVRA	W0T602	19	8.68	1.811
MIGALFLFWVFFTHLASA	W0T6E2	18	6.49	1.811
MMVLTQLILCFCSWTIAVSA	W0TH45	20	5.27	1.830
MVLRLDFRLITLLILIISVGVVHISS	W0TJN4	26	9.37	1.831
MVALRLLISGALLSVASA	W0T7Q2	18	9.50	1.872
MRFLILLSIFAVAIRG	W0T7W2	16	12.00	1.875
MWFALFQVLLLMGIPLVDC	W0TE01	19	3.80	1.879
MLAILCLLFTLNPFWAAA	W0TDM7	18	5.27	1.889
MKLPLFLLFLIASVYA	W0TCP8	16	8.34	1.950
MKVALILALASTALA	W0TDI8	15	8.50	1.960
MVSVLVFLVLSVTQFARLVFS	W0TFR8	21	9.50	1.971
MMSLAKFITFFLGLLLFLLVSLRVLS	W0TFP9	26	11.00	2.000
MMLSINQALSVLLFLVSVCFA	W0T4Y9	21	5.27	2.010
MNLQGIVTVLLCLISFALA	W0THC0	19	5.27	2.016
MIKILLSLVFLLLAQLVHS	W0TBV0	19	8.52	2.016
MRIIVLFLAFLQCFSGCL	W0TCX7	18	7.83	2.017

MSVLIRVLVLVQLLARLLLVQA	W0T333	22	12.00	2.027
MFPLRFVLFAILQLIFVSA	A0A2Z4HVV7	19	9.50	2.047
MLLQFLVLLFAGLCNA	W0T9G6	16	5.27	2.075
MEVFLLLCLLFSATVTA	W0TA76	17	4.00	2.076
MSKVVAVPMVLLLLLSLVNA	W0T627	21	8.50	2.114
MWFVIFQVLLVLGISFVDC	W0T4Y3	19	3.80	2.153
MVKLLSLLSFALLPFVLA	W0T4H9	18	8.50	2.167
MKLLLHIFSVLSFLAAIVFA	W0TCA0	20	8.52	2.170
MILTCWIVNIWISIVLVSA	W0T979	19	5.27	2.174
MYSVILLLMTIGVVNVLA	W0TBM8	18	5.27	2.217
MLFPIAFLACVVCASG	W0T6Y4	16	5.27	2.225
MKLFLQSIILCFCALIAVA	W0T3P3	19	7.82	2.242
MIFALIPLLTVFFCSFCAA	W0TE17	19	5.27	2.368
MRVLLLLFLYVFGVIA	W0THY4	16	8.50	2.450

Table S4. Secretion Signals identified *in silico*. Amino acid sequence, protein identifier, length, isoelectric point of the signal and hydrophobicity.

Script S1. Python script used for visualisation of the SS 3D space. This script generates a three-dimensional visualisation for SS properties. Each SS is represented as a point within a normalised feature space defined by length, isoelectric point, and hydrophobicity. A semi-transparent cube is used to bound the parameter space, and all proteins are plotted to illustrate their overall distribution. The selected subset of SSs is highlighted to facilitate comparison with the full dataset.

```
import matplotlib.pyplot as plt
from mpl_toolkits.mplot3d.art3d import Poly3DCollection

plt.rcParams['font.family'] = 'Tw Cen MT'

# Create a figure and a 3D axis
fig = plt.figure()
ax = fig.add_subplot(111, projection='3d')

# Plot the cube
cube_vertices = [
    [0, 0, 0],
    [1, 0, 0],
    [1, 1, 0],
    [0, 1, 0],
    [0, 0, 1],
    [1, 0, 1],
    [1, 1, 1],
    [0, 1, 1]
]

cube_faces = [
    [0, 1, 2, 3],  # Bottom face
    [4, 5, 6, 7],  # Top face
    [0, 1, 5, 4],  # Front face
    [1, 2, 6, 5],  # Right face
    [2, 3, 7, 6],  # Back face
```

```
    [3, 0, 4, 7]   # Left face
    ]

transparency = 0.2  # Set the transparency value (0.0 - 1.0)

for face in cube_faces:
    vertices = [cube_vertices[vertex] for vertex in face]
    poly = Poly3DCollection([vertices], alpha=0.1)
    poly.set_facecolor('yellow')
    ax.add_collection3d(poly)

# Define the list of dots with names
dot_list = [
 {'coords': [0.9375, 0.879289940828402, 0], 'name': '1'},
 #{'coords': [0.875, 0.899408284023669, 0.205539358600584], 'name': 'SUN4'}
 {'coords': [0.25, 0.564497041420118, 0.233560090702947], 'name': '3'},
 {'coords': [0.5, 0.257988165680473, 0.259740259740261], 'name': '4'},
 {'coords': [0.6875, 0.879289940828402, 0.267755102040816], 'name': '5'},
 {'coords': [0.625, 0.791715976331361, 0.272108843537416], 'name': '6'},
 {'coords': [0.375, 0.564497041420118, 0.279591836734694], 'name': '7'},
 #{'coords': [0.625, 0, 0.287414965986396], 'name': 'GMC1'},
 {'coords': [1, 0.34792899408284, 0.291156462585033], 'name': '9'},
 {'coords': [0.5625, 0.518343195266272, 0.292812777284829], 'name': '10'},
 #{'coords':  [0.1875,  0.997633136094675,  0.292917166866747],  'name':
'30437'},
 {'coords': [0.6875, 0.275739644970414, 0.313469387755102], 'name': '12'},
 {'coords': [0.25, 0.701775147928994, 0.324263038548751], 'name': '13'},
 {'coords': [0.25, 0.76094674556213, 0.324263038548751], 'name': '14'},
 {'coords': [0.25, 0.585798816568047, 0.335600907029478], 'name': '15'},
 {'coords': [0.4375, 0.52189349112426, 0.347910592808551], 'name': '16'},
 {'coords': [0.5, 0.67810650887574, 0.348794063079775], 'name': '17'},
 {'coords': [0.9375, 0.764497041420118, 0.353272343420126], 'name': '18'},
 {'coords': [0.625, 0.899408284023669, 0.370748299319727], 'name': '19'},
```

{'coords': [0.625, 0.275739644970414, 0.372448979591837], 'name': '20'},
{'coords': [0.875, 0.701775147928994, 0.380466472303208], 'name': '21'},
{'coords': [0.5625, 1, 0.383318544809229], 'name': '22'},
{'coords': [0.3125, 0.583431952662722, 0.384532760472608], 'name': '23'},
{'coords': [0.5625, 0.564497041420118, 0.395740905057673], 'name': '24'},
{'coords': [0.375, 0.679289940828402, 0.395918367346939], 'name': '25'},
{'coords': [0.3125, 0.518343195266272, 0.414607948442531], 'name': '26'},
{'coords': [0.3125, 0.257988165680473, 0.416756176154653], 'name': '27'},
{'coords': [0.3125, 0.564497041420118, 0.416756176154653], 'name': '28'},
{'coords': [0.625, 0.0757396449704142, 0.420068027210898], 'name': '29'},
{'coords': [0.5625, 0.76094674556213, 0.424134871339837], 'name': '30'},
{'coords': [0.3125, 0.76094674556213, 0.427497314715347], 'name': '31'},
{'coords': [0.375, 0.796449704142012, 0.428571428571429], 'name': '32'},
{'coords': [0.5625, 0.23905325443787, 0.429458740017755], 'name': '33'},
{'coords': [0.75, 0.654437869822485, 0.434850863422286], 'name': '34'},
{'coords': [0.4375, 0.635502958579882, 0.435374149659878], 'name': '35'},
{'coords': [0.25, 0.583431952662722, 0.444444444444449], 'name': '36'},
{'coords': [0.625, 0.67810650887574, 0.445578231292531], 'name': '37'},
{'coords': [0.5625, 0.345562130177515, 0.447204968944082], 'name': '38'},
{'coords': [0.6875, 0.998816568047337, 0.448979591836735], 'name': '39'},
{'coords': [0.375, 0.583431952662722, 0.451020408163265], 'name': '40'},
{'coords': [0.25, 0.76094674556213, 0.451247165532898], 'name': '41'},
{'coords': [0.1875, 0.859171597633136, 0.456182472989184], 'name': '42'},
{'coords': [0.4375, 0.802366863905325, 0.460641399416898], 'name': '43'},
{'coords': [0.3125, 0.518343195266272, 0.470461868958122], 'name': '44'},
{'coords': [0.5, 0.382248520710059, 0.473098330241184], 'name': '45'},
{'coords': [0.5625, 0.618934911242604, 0.473824312333633], 'name': '46'},
{'coords': [0.3125, 0.687573964497042, 0.474758324382367], 'name': '47'},
{'coords': [0.25, 0.679289940828402, 0.478458049886612], 'name': '48'},
{'coords': [0.6875, 0.585798816568047, 0.48], 'name': '49'},
{'coords': [0.25, 0.583431952662722, 0.494331065759633], 'name': '50'},
{'coords': [0.4375, 0.76094674556213, 0.505344995140898], 'name': '51'},
{'coords': [0.5, 0.583431952662722, 0.515769944341388], 'name': '52'},

{'coords': [0.375, 0.500591715976331, 0.518367346938776], 'name': '53'},

{'coords': [0.125, 0.564497041420118, 0.528061224489796], 'name': '54'},

{'coords': [0.375, 0.565680473372781, 0.528571428571428], 'name': '55'},

#{'coords': [0.25, 0.0272189349112426, 0.532879818594122], 'name': 'ECM38'},

#{'coords': [0.625, 0.52189349112426, 0.534013605442163], 'name': 'MSB2'},

{'coords': [0.4375, 0.701775147928994, 0.534499514091347], 'name': '58'},

{'coords': [0.375, 0.583431952662722, 0.542857142857143], 'name': '59'},

{'coords': [0.3125, 0.879289940828402, 0.543501611170776], 'name': '60'},

#{'coords': [0.4375, 0.553846153846154, 0.54616132167151], 'name': 'KLMX_10834'},

{'coords': [0.3125, 0.583431952662722, 0.547798066595061], 'name': '62'},

{'coords': [0.25, 0.725443786982248, 0.548752834467102], 'name': '63'},

{'coords': [0.1875, 0.583431952662722, 0.552220888355347], 'name': '64'},

{'coords': [0.6875, 0.583431952662722, 0.553469387755102], 'name': '65'},

{'coords': [0.5, 0.257988165680473, 0.556586270872], 'name': '66'},

#{'coords': [0.4375, 0.524260355029586, 0.557823129251714], 'name': 'KRE1'},

{'coords': [0.25, 0.583431952662722, 0.560090702947837], 'name': '68'},

{'coords': [0.375, 0.201183431952663, 0.56530612244898], 'name': '69'},

{'coords': [0.25, 0.736094674556213, 0.566893424036286], 'name': '70'},

{'coords': [0.4375, 1, 0.571428571428571], 'name': '71'},

{'coords': [0.0625, 0.564497041420118, 0.579591836734694], 'name': '72'},

{'coords': [0.4375, 0.280473372781065, 0.583090379008735], 'name': '73'},

{'coords': [0.3125, 0.344378698224852, 0.584317937701388], 'name': '74'},

{'coords': [0.3125, 0.583431952662722, 0.586466165413551], 'name': '75'},

{'coords': [0.25, 0.346745562130178, 0.591836734693878], 'name': '76'},

{'coords': [0.25, 0.202366863905326, 0.594104308390041], 'name': '77'},

{'coords': [0.4375, 0.583431952662722, 0.594752186588939], 'name': '78'},

{'coords': [0.375, 0.518343195266272, 0.597959183673469], 'name': '79'},

{'coords': [0.3125, 0.76094674556213, 0.599355531686367], 'name': '80'},

{'coords': [0.4375, 0.0260355029585799, 0.60058309037902], 'name': '81'},

{'coords': [0.5, 0.345562130177515, 0.601113172541755], 'name': '82'},

{'coords': [0.375, 0.518343195266272, 0.602040816326531], 'name': '83'},

{'coords': [0.4375, 0.201183431952663, 0.604470359572408], 'name': '84'},

{'coords': [0.4375, 0.583431952662722, 0.606413994169102], 'name': '85'},
{'coords': [0.4375, 0.583431952662722, 0.606413994169102], 'name': '86'},
{'coords': [0.1875, 0.0508875739644971, 0.607442977190857], 'name': '87'},
{'coords': [0.375, 0.736094674556213, 0.608163265306122], 'name': '88'},
{'coords': [0.25, 0.629585798816568, 0.60997732426302], 'name': '89'},
{'coords': [0.4375, 0.369230769230769, 0.61030126336249], 'name': '90'},
{'coords': [0.25, 0.202366863905326, 0.614512471655347], 'name': '91'},
{'coords': [0.3125, 0.345562130177515, 0.618689581095592], 'name': '92'},
{'coords': [0.4375, 0.52189349112426, 0.620019436345959], 'name': '93'},
{'coords': [0.5, 0.0946745562130178, 0.625231910946204], 'name': '94'},
{'coords': [0.25, 0.583431952662722, 0.628117913832204], 'name': '95'},
{'coords': [0.375, 0.701775147928994, 0.628571428571429], 'name': '96'},
{'coords': [0.3125, 0.23905325443787, 0.629430719656286], 'name': '97'},
{'coords': [0.25, 0.201183431952663, 0.630385487528327], 'name': '98'},
{'coords': [0.375, 0.76094674556213, 0.63469387755102], 'name': '99'},
{'coords': [0.4375, 0.201183431952663, 0.635568513119551], 'name': '100'},
{'coords': [0, 0.583431952662722, 0.641399416909633], 'name': '101'},
{'coords': [0.375, 0.202366863905326, 0.646938775510204], 'name': '102'},
{'coords': [0.1875, 0.67810650887574, 0.653061224489796], 'name': '103'},
{'coords': [0.25, 0.583431952662722, 0.653061224489796], 'name': '104'},
{'coords': [0.4375, 0.201183431952663, 0.653061224489796], 'name': '105'},
{'coords': [0.5625, 0.515976331360947, 0.656610470275061], 'name': '106'},
{'coords': [0.4375, 0.583431952662722, 0.656948493683184], 'name': '107'},
{'coords': [0.4375, 0.701775147928994, 0.656948493683184], 'name': '108'},
{'coords': [0.1875, 0.201183431952663, 0.660264105642245], 'name': '109'},
{'coords': [0.25, 0.583431952662722, 0.664399092970531], 'name': '110'},
{'coords': [0.25, 0.67810650887574, 0.664399092970531], 'name': '111'},
{'coords': [0.4375, 0.202366863905326, 0.666666666666653], 'name': '112'},
{'coords': [0.375, 0.492307692307692, 0.673469387755102], 'name': '113'},
{'coords': [0.125, 0.583431952662722, 0.678571428571429], 'name': '114'},
#{'coords': [0.6875, 0.201183431952663, 0.685714285714286], 'name': 'DSE4'},
{'coords': [0.4375, 0.701775147928994, 0.686103012633633], 'name': '116'},
{'coords': [0.3125, 0.52189349112426, 0.689581095596122], 'name': '117'},

{'coords': [0.25, 0.52189349112426, 0.696145124716571], 'name': '118'},
{'coords': [0.4375, 0.201183431952663, 0.69970845481049], 'name': '119'},
{'coords': [0.4375, 0.583431952662722, 0.69970845481049], 'name': '120'},
{'coords': [0.125, 0.524260355029586, 0.701530612244898], 'name': '121'},
{'coords': [0.3125, 0.583431952662722, 0.706766917293224], 'name': '122'},
{'coords': [0.5, 0.201183431952663, 0.708719851576979], 'name': '123'},
{'coords': [0.3125, 0.524260355029586, 0.717508055853918], 'name': '124'},
{'coords': [0.375, 0.52189349112426, 0.718367346938775], 'name': '125'},
{'coords': [0.375, 0.52189349112426, 0.724489795918367], 'name': '126'},
{'coords': [0.1875, 0.201183431952663, 0.725090036014408], 'name': '127'},
{'coords': [0.3125, 0.202366863905326, 0.732545649838898], 'name': '128'},
{'coords': [0.375, 0.201183431952663, 0.73265306122449], 'name': '129'},
{'coords': [0.375, 0.201183431952663, 0.73265306122449], 'name': '130'},
{'coords': [0.125, 0.583431952662722, 0.73469387755102], 'name': '131'},
{'coords': [0.3125, 0.643786982248521, 0.73469387755102], 'name': '132'},
{'coords': [0.3125, 0.604733727810651, 0.738990332975306], 'name': '133'},
{'coords': [0.25, 0.345562130177515, 0.739229024943306], 'name': '134'},
{'coords': [0.375, 0.201183431952663, 0.746938775510204], 'name': '135'},
{'coords': [0.75, 0.686390532544379, 0.747252747252735], 'name': '136'},
{'coords': [0.25, 0.701775147928994, 0.764172335600898], 'name': '137'},
#{'coords': [0.125, 0.997633136094675, 0.76530612244898], 'name': 'ROT1'}
{'coords': [0.3125, 0.0272189349112426, 0.766917293233102], 'name': '139'},
{'coords': [0.25, 0.201183431952663, 0.770975056689347], 'name': '140'},
{'coords': [0.125, 0.564497041420118, 0.795918367346939], 'name': '141'},
{'coords': [0.0625, 0.583431952662722, 0.8], 'name': '142'},
{'coords': [0.4375, 0.701775147928994, 0.804664723032082], 'name': '143'},
#{'coords': [0.75, 0.879289940828402, 0.816326530612245], 'name': 'ALG11'},
{'coords': [0.4375, 0.201183431952663, 0.820213799805633], 'name': '145'},
{'coords': [0.3125, 0.201183431952663, 0.822771213748653], 'name': '146'},
{'coords': [0.3125, 0.585798816568047, 0.822771213748653], 'name': '147'},
{'coords': [0.25, 0.50414201183432, 0.823129251700694], 'name': '148'},
{'coords': [0.5, 0.997633136094675, 0.827458256029673], 'name': '149'},
{'coords': [0.3125, 0.701775147928994, 0.835660580021469], 'name': '150'},

```
{'coords': [0.125, 0.201183431952663, 0.846938775510204], 'name': '151'},
{'coords': [0.1875, 0.0508875739644971, 0.847539015606245], 'name': '152'},
{'coords': [0.4375, 0.583431952662722, 0.862973760932939], 'name': '153'},
{'coords': [0.3125, 0.0272189349112426, 0.878625134264245], 'name': '154'},
{'coords': [0.25, 0.583431952662722, 0.884353741496612], 'name': '155'},
{'coords': [0.375, 0.585798816568047, 0.885714285714286], 'name': '156'},
{'coords': [0.3125, 0.201183431952663, 0.887218045112775], 'name': '157'},
{'coords': [0.25, 0.201183431952663, 0.904761904761918], 'name': '158'},
#{'coords': [0.125, 0.201183431952663, 0.908163265306122], 'name': 'GTB1'},
{'coords': [0.3125, 0.502958579881657, 0.915145005370571], 'name': '160'},
{'coords': [0.3125, 0.201183431952663, 0.966702470461878], 'name': '161'},
{'coords': [0.125, 0.583431952662722, 1], 'name': '162'}
]


selected_list = [
{'coords': [0.25, 0.0272189349112426, 0.532879818594122], 'name': 'ECM38'},
{'coords': [0.1875, 0.997633136094675, 0.292917166866747], 'name':
'KLMX_30437'},
{'coords': [0.125, 0.201183431952663, 0.908163265306122], 'name': 'GTB1'},
{'coords': [0.125, 0.997633136094675, 0.76530612244898], 'name': 'ROT1'},
{'coords': [0.625, 0, 0.287414965986396], 'name': 'GMC1'},
{'coords': [0.875, 0.899408284023669, 0.205539358600584], 'name': 'SUN4'},
{'coords': [0.6875, 0.201183431952663, 0.685714285714286], 'name': 'DSE4'},
{'coords': [0.75, 0.879289940828402, 0.816326530612245], 'name': 'ALG11'},
{'coords': [0.625, 0.52189349112426, 0.534013605442163], 'name': 'MSB2'},
{'coords': [0.4375, 0.553846153846154, 0.54616132167151], 'name':
'KLMX_10834'},
{'coords': [0.4375, 0.524260355029586, 0.557823129251714], 'name': 'KRE1'}
]


dot_size = 5  # Set the size of the dots


for dot in dot_list:
```

```
    coords = dot['coords']
    name = dot['name']
    ax.scatter(*coords, color='green', s=dot_size, alpha=0.3)
      ax.text(*coords,  name,  fontsize=10,  color='black',  ha='center',
va='bottom', alpha=0.3)

for dot in selected_list:
    coords = dot['coords']
    name = dot['name']
    ax.scatter(*coords, color='blue', s=dot_size)
      ax.text(*coords,  name,  fontsize=10,  color='black',  ha='center',
va='bottom')

# Set the limits of the plot
ax.set_xlim([1, 0])
ax.set_ylim([0, 1])
ax.set_zlim([0, 1])

# Set the aspect ratio of the plot
ax.set_box_aspect([1, 1, 1])

# Set the axis labels
ax.set_xlabel('Length', fontsize=15)
ax.set_ylabel('Isopoint', fontsize=15)
ax.set_zlabel('Hydrophobicity', fontsize=15)

# Show the plot
plt.show()
```

Script S2. Python script used for analysis of the SSs. This script performs data analysis and visualisation of SS features: length, isoelectric point, and hydrophobicity (GRAVY). Distributions of each feature are visualised using violin plots combined with swarm plots, as well as histograms with kernel density estimation. Pairwise relationships and correlations between variables are assessed using a correlation heatmap and pairplot. Principal component analysis (PCA) is applied to standardised features, and the resulting two-dimensional projection is visualised to assess overall variation. Selected SSs are highlighted within the PCA space to facilitate comparison with the full dataset.

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_excel(r"file_path/file.xlsx")

features = ['Length', 'Isoelectric point', 'GRAVY']
titles = {
    'Length': 'Length Distribution',
    'Isoelectric point': 'Isoelectric point Distribution',
    'GRAVY': 'Hydrophobicity Distribution'
}

plt.figure(figsize=(8, 14))  # Taller figure for better spacing

colors = ["#6D89B6", "#88C897", "#DF6CC6"]

for i, feature in enumerate(features, 1):
    plt.subplot(3, 1, i)

    # Create dummy x so violin has space to expand
        sns.violinplot(x=[""]   *   len(df),   y=df[feature],   inner='box',
scale='width', color=colors[i-1])

    # Add actual data points
```

```
     sns.swarmplot(x=[""] * len(df), y=df[feature], color='k', alpha=0.3,
size=3)


    plt.title(titles[feature], fontsize=14)
    plt.ylabel(feature, fontsize=12)
    plt.xlabel("")  # Remove x label
    plt.xticks([])  # Remove tick label on x
    plt.grid(axis='y', linestyle='--', alpha=0.3)
    sns.despine(left=True, bottom=True)

# Add spacing between subplots
plt.subplots_adjust(hspace=0.4)
plt.tight_layout()
plt.show()

# Histogram + KDE for each variable
for col in ['Length', 'Isoelectric point', 'GRAVY']:
    plt.figure()
    sns.histplot(df[col], kde=True)
    plt.title(f'Distribution of {col}')
    plt.show()

corr = df[['Length', 'Isoelectric point', 'GRAVY']].corr()

# Heatmap
plt.figure(figsize=(6, 4))
sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()

sns.pairplot(df[['Length', 'Isoelectric point', 'GRAVY']])
plt.show()
```

```
# PCA

from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

features = ['Length', 'Isoelectric point', 'GRAVY']
X = df[features]
X_scaled = StandardScaler().fit_transform(X)

pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)

# Add PCA results back to your dataframe
df['PC1'] = X_pca[:, 0]
df['PC2'] = X_pca[:, 1]

import matplotlib.pyplot as plt

# List of identifiers to highlight
highlight_ids = {
    'W0TFP9', 'W0T4U4', 'W0T4A4', 'W0T2A3', 'W0T6Y4',
    'W0TH48', 'W0TAK9', 'W0T7W2', 'W0TI90', 'W0T5X0', 'W0T8J1'
}

# Base scatter plot for all points
plt.figure(figsize=(12, 8))
plt.scatter(df['PC1'], df['PC2'], alpha=0.4, label='All SSs', color='gray')

# Highlight selected identifiers
highlight_df = df[df['Protein Code'].isin(highlight_ids)]
plt.scatter(highlight_df['PC1'],     highlight_df['PC2'],     color='red',
label='SSs selected')
```

```
# Add text labels for highlighted points
for i, row in highlight_df.iterrows():
    plt.text(row['PC1'] + 0.02, row['PC2'], row['Protein Code'], fontsize=8,
color='red')


# Final touches
plt.xlabel('PC1')
plt.ylabel('PC2')
plt.title('PCA of Secretion Signal Features')
plt.grid(True)
plt.legend()
plt.tight_layout()
plt.show()
```

Script S3. Python script used for analysis of SSs performance across assays. This script performs statistical comparison and visualisation of normalised measurements. Raw values are first normalised within each group, and the resulting data are subjected to pairwise comparisons using Tukey's honestly significant difference (HSD) test. A compact letter display is generated to summarise statistically significant differences between groups. Results are visualised as bar plots showing group means with standard error, overlaid with individual data points. Statistical groupings are indicated above each bar, allowing clear interpretation of differences among groups while preserving the distribution of replicate measurements.

```
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from collections import defaultdict


plt.rcParams['font.family'] = 'Tw Cen MT'


file_path = r"file.xlsx"
df = pd.read_excel(file_path)
```

```
# Greek replacement and formatting
greek_map = {'α': r'$\alpha$'}

if 'Signal' in df.columns:
    df['Signal'] = df['Signal'].apply(lambda s: str(s).replace(' ', '\n'))
    df['Signal'] = df['Signal'].apply(
        lambda s: ''.join(greek_map.get(ch, ch) for ch in s)
    )

# Capture original order from Excel (Order of appearance)
strain_order = df[df.columns[0]].unique().tolist()

# Normalize
strains = {}
for strain, g in df.groupby(df.columns[0], sort=False):
    a = g.iloc[:, 1].astype(float).to_numpy()
    o = g.iloc[:, 2].astype(float).to_numpy()
    strains[strain] = a / o

# Flatten for Tukey
values = []
labels = []
for strain, arr in strains.items():
    for v in arr:
        values.append(v)
        labels.append(strain)

values = np.array(values)
labels = np.array(labels)

# Tukey
tukey = pairwise_tukeyhsd(endog=values, groups=labels, alpha=0.05)
print(tukey.summary())
```

```
# ---- Compact letter assignment ----
groups = sorted(strains.keys())
letters = {g: set() for g in groups}

# Build significance matrix
sig = defaultdict(dict)
for row in tukey._results_table.data[1:]:
    g1, g2, meandiff, p, lower, upper, reject = row
    sig[g1][g2] = reject
    sig[g2][g1] = reject

# Assign letters iteratively
sorted_groups_for_stats = sorted(groups, key=lambda g: strains[g].mean())
current_letter = 'A'

for g in sorted_groups_for_stats:
    # try to fit into an existing letter
    placed = False
    for letter in letters[g] or []:
          if all(not sig[g].get(h, False) for h in groups if letter in
letters[h]):
            letters[g].add(letter)
            placed = True
            break
    if placed: continue

    # otherwise assign new letter
    letters[g].add(current_letter)
    # give same letter to groups not significantly different
    for h in sorted_groups_for_stats:
        if not sig[g].get(h, False):
            letters[h].add(current_letter)
```

```
    current_letter = chr(ord(current_letter) + 1)


# ---- PREPARE DATA FOR PLOTTING ----


groups_sorted = strain_order


# Extract means and SEs in this specific order

means = [np.mean(strains[g]) for g in groups_sorted]

ses = [np.std(strains[g], ddof=1) / np.sqrt(len(strains[g])) for g in
groups_sorted]


# Plot ----------------------------------------------------

x = np.arange(len(groups_sorted))

plt.figure(figsize=(10, 6))


cmap = plt.colormaps.get_cmap("tab20")  # Use 'tab20' for more colors


bars = plt.bar(x, means, yerr=ses, alpha=0.6, width=0.5, capsize=4,

               color = cmap(x / len(groups_sorted)))


# jittered points

for i, g in enumerate(groups_sorted):

    arr = strains[g]

    jitter = (np.random.rand(len(arr)) - 0.5) * 0.15

    plt.scatter(np.repeat(i, len(arr)) + jitter, arr, color='k', s=30)


#add compact letters

for i, g in enumerate(groups_sorted):

    letter = ''.join(sorted(letters[g]))

    # Adjust height of letter above the error bar

    y = means[i] + ses[i] + 0.05 * max(means)

    plt.text(i, y, letter, ha='center', fontsize=16)
```

```
plt.xticks(x, groups_sorted, rotation=45, fontsize=12)
plt.yticks(fontsize=20)
plt.ylabel(fontsize=20)
plt.title(fontsize=40)

plt.tight_layout()
plt.show()
```

Chapter 4

Evolving an improved strain of *Kluyveromyces marxianus* for use in a dairy biorefinery

Abstract

Whey permeate, a by-product of dairy processing, is a promising feedstock for biotechnology applications due to its high lactose content and widespread availability. Carbery Food Ingredients, a food ingredients and cheese manufacturer based in Ireland, pioneered the conversion of whey permeate into bioethanol by developing an industrial whey-to-ethanol plant in 1978 using the yeast *K. marxianus*. However, there is a need for a more efficient yeast strain capable of faster fermentation turnover and greater resilience against fluctuating process conditions. Regulatory frameworks and consumer preferences limit the use of Genetically Modified Organisms, which might otherwise address these limitations. Adaptive Laboratory Evolution offers a powerful approach to strain improvement by harnessing natural selection under controlled conditions. In this study, an Adaptive Laboratory Evolution strategy was applied to *K. marxianus* for 600 generations, utilising ethanol stress as a selective pressure to drive the evolution of general strain robustness. Yeast variants that exhibited improved tolerance were identified and characterised in the laboratory. The most promising variant was selected for scale-up trials at Carbery's facilities, and demonstrated superior fermentation performance and stability, confirming that selection for ethanol tolerance successfully conferred enhanced robustness suitable for an industrial environment. This work highlights the potential of Adaptive Laboratory Evolution as an effective non-Genetically Modifying strategy to improve industrial yeast performance under challenging fermentation conditions, thereby supporting the valorisation of dairy waste streams for renewable fuel production.

Introduction

The dairy industry faces many sustainability issues (Capper et al., 2009; Finnegan & Goggins, 2021; Guzmán-Luna et al., 2022). It is a significant global contributor to greenhouse gas emissions, primarily methane from enteric fermentation in cows and nitrous oxide from manure and fertiliser use. Dairy farming also requires substantial amounts of water for irrigation, animal drinking, and processing of goods. Moreover, farms typically occupy a large amount of agricultural land and have a negative impact on biodiversity. The European Green Deal (European Commission, 2019) sets the European Union (EU) to become climate-neutral by 2050. Hence, the dairy sector is under enormous pressure in order to become more sustainable. Biorefineries stand out as an elegant and circular solution to achieve sustainability goals. A dairy biorefinery converts dairy manufacturing side-streams into a range of bio-based products including biofuels, platform chemicals, and nutraceuticals, through the application of advanced membrane filtration, enzymatic hydrolysis, and microbial fermentation. This systematic approach maximises the economic value of milk while significantly reducing the environmental footprint of the dairy industry, aligning with the principles of a circular bioeconomy (Asunis et al., 2020; Mukherjee et al., 2023; Zotta et al., 2020).

Carbery Food Ingredients (Carbery), an Irish dairy company located in Cork, is a major producer of cheese, flavours, proteins and bioethanol. The Carbery production facility exemplifies a high-efficiency dairy biorefinery that integrates primary manufacturing with advanced waste valorisation to achieve a circular bioeconomy (Fig. 1a). The facility utilises a cascading approach to fractionate milk into diverse high-value streams. Following primary cheese production, the resulting whey is processed through a hierarchy of membrane filtration technologies to recover functional whey protein concentrates and isolates, alongside specialised hydrolysed proteins for clinical and performance nutrition. The remaining lactose-rich by-product, WP, is diverted to a fermentation stream, where it is converted into bioethanol and yeast extract, significantly reducing the organic load of the plant's side-streams. To achieve full resource circularity, the final waste effluent undergoes anaerobic digestion to produce biogas, which supplies 9% of the natural gas required

for the facility's combined heat and power (CHP) plant. This loop is closed by the recovery of an organic fertiliser and treated water, which are returned to local farms to support the grass-based dairy production system, effectively minimising environmental discharge and maximising resource productivity.

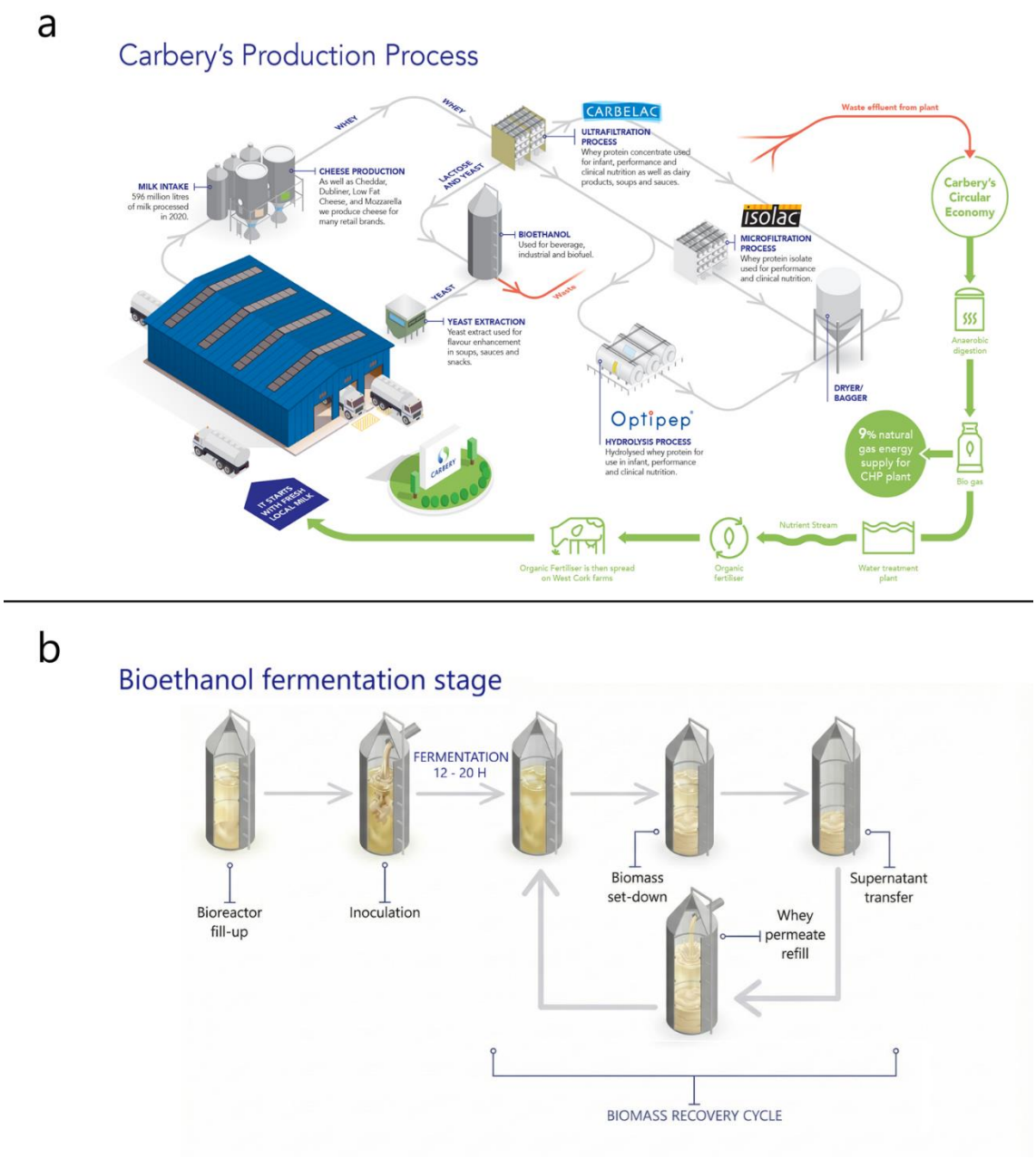


Fig. 1. The Carbery biorefinery approach. a) Diagram showing Carbery's circular production process. Obtained from Carbery's website (https://www.carbery.com/nutrition/product-range/). **b)** Detail on the bioethanol production process and biomass recovery (re-pitching) cycle.

In particular, the alcoholic fermentation process at Carbery runs 11 cylindroconical vessels for the production of bioethanol, with an approximate capacity of 1200 hL

per vessel. The fermentation process relies on the dairy yeast *K. marxianus* to transform lactose into bioethanol in a cyclic manner (Fig. 1b). The bioprocess starts by filling up a fermentation vessel with WP. The fermentation, run in batch mode, begins when the WP is inoculated with yeast, and during growth, compressed air is continuously sparged at a low rate to supply oxygen for lipid synthesis and achieve cell suspension. The batch lasts 12 to 20 hours, according to the initial concentration of yeast cells and lactose. Fermentation continues until lactose exhaustion, at this stage the air sparger is stopped, and the biomass is left to settle. Once separated, the supernatant is diverted for distillation, and the remaining yeast is re-used for fermentation by refilling the vessel with fresh WP. This cycle is repeated until yeast exhaustion (typically 4 to 5 cycles). Ethanol titres at the end of the process are on average in the range of 2.5 to 4.2% (v/v) depending on the initial lactose concentration (Pesta et al. 2007; Ling 2008; Hamilton 1998; Thiele 2005) according to the stoichiometry of lactose alcoholic fermentation (Equation 1).

$$C_{12}H_{22}O_{11} + H_2O \longrightarrow 4C_2H_6O + 4CO_2$$

Equation 1: Stoichiometry of lactose alcoholic fermentation.

K. marxianus is a yeast belonging to the Saccharomycetaceae family that is particularly attractive for the production of bioethanol from dairy side streams. Unlike *S. cerevisiae*, *K. marxianus* natively metabolises lactose (Varela et al., 2019). Additionally, its rapid growth rate and thermotolerance provide further advantages for industrial-scale fermentations, including shortened process times, reduced cooling costs, and enhanced resistance to microbial contamination, which are critical factors when handling dairy effluents. Moreover, its metabolic versatility and expanding engineering toolbox make *K. marxianus* a highly adaptable microbial cell factory for the development of sustainable dairy-based biorefineries. This ability enables direct valorisation of dairy wastes (Guimarães et al., 2010).

This bioprocess, although simple, faces different challenges mainly because of the great scale at which is performed. The main challenge lays on WP variability along the year: milk characteristics are heterogeneous varying in protein, fat and lactose contents depending on the season (da Silveira Garcia et al., 2024). Moreover, during

peak milk production, reverse osmosis (RO) is applied after the ultrafiltration step to reduce WP volume (Fig. 1a), generating a more concentrated lactose stream known as high-gravity permeate typically above 80 g L^{-1} (Pasotti et al., 2017). While *S. cerevisiae* is known for high ethanol tolerance, *K. marxianus* strains typically can only withstand concentrations in the 5 to 7 % v/v range (Costa et al., 2014; da Silveira et al., 2020; Saini et al., 2017). Consequently, when lactose-rich streams (>100 g L^{-1}) are fermented, ethanol concentrations can surpass the native tolerance limit of *K. marxianus*. Moreover, because the Carbery bioethanol production is performed as a re-pitched batch, issues such as bacterial contamination, higher concentrations of salts and acids or yeast senescence are common (Pratt et al., 2003; Tesfaw, 2023; Wauters et al., 2021; Yaakoub et al., 2022). For all these reasons, the composition and strength of WP can vary depending on production cycles and processing needs, while the yeast must consistently maintain performance under a range of substrate and stress conditions. This scenario underscores the value of developing robust *K. marxianus* strains capable of sustaining efficient fermentation across fluctuating industrial environments.

Given the multiple industrial challenges *K. marxianus* faces in a dairy biorefinery, there is a critical need to develop an improved *K. marxianus* strain with enhanced robustness. Ethanol tolerance involves complex, polygenic interactions that span cell membrane composition, protein folding stability, and energy metabolism (de Moura Ferreira et al., 2022). For this reason, ethanol tolerance was selected as the primary selective pressure for this study. By evolving a strain specifically to withstand higher ethanol concentrations, which can currently exceed the native tolerance limit of ~5% (v/v), we aim to simultaneously confer cross-protection against other industrial stressors. This approach seeks to overcome the current limitations of the re-pitched batch process, where yeast senescence and environmental variability currently restrict the system to only 4 to 5 cycles before exhaustion.

ALE is a powerful experimental strategy that harnesses the principles of natural selection to improve microbial performance under specific selective conditions. By cultivating microorganisms over many generations in challenging environments, ALE promotes the accumulation of beneficial mutations that enhance traits such as speed

of growth, tolerance to inhibitors, or product yield (Dragosits & Mattanovich, 2013; Mavrommati et al., 2022). ALE is a highly effective tool to complement rational metabolic engineering. While genetic modifications can target specific pathways, ALE provides an unbiased and systematic route to uncover novel adaptations that involve multi-genic factors and might otherwise remain elusive. This makes it especially valuable for optimising microbial hosts to thrive on complex feedstocks, withstand harsh process conditions, and achieve higher productivity, thereby accelerating the development of sustainable and economically viable bioprocesses. However, and due to the "black box" nature of ALE, scale-up of these strains must involve careful, stepwise validation since unwanted features can appear during the process (Lee & Kim, 2020).

In this work, we first investigated the ethanol tolerance of the Carbery strain of *K. marxianus* and performed an ALE based strategy to improve its ethanol tolerance. After 600 generations, we isolated strains with increased ethanol tolerance, and characterised their ability to withstand ethanol as well as examining any possible trade-offs that could appear during the ALE under industrial conditions. Finally, we brought the strain to the industrial plant and analysed its behaviour during operation conditions. We confirmed that the evolved strain has increased ethanol tolerance and therefore is more robust when operating in industrial environments. This work establishes ALE as a vital tool for developing resilient, non-GMO yeast strains capable of maximising the economic and environmental value of dairy side-streams, while underscoring the critical importance of validating these strains at industrial scale.

Material and methods

Strains and growth media

The diploid dairy-adapted *K. marxianus* CBS 397 (WT) strain (Ortiz-Merino et al., 2018) was purchased from The Westerdijk Fungal Biodiversity Institute (Utrecht, The Netherlands) in lyophilised form. Subsequently, it was cultured in YPLac medium (yeast extract, 10 g L^{-1}; peptone, 20 g L^{-1}; lactose, 20 g L^{-1}) for 20 to 24 hours and stocked in sterile 60% (v/v) glycerol to a final concentration of 24% (v/v) and 1 mL aliquots of this mixture were stocked frozen at -80 °C. For the preparation of experiments, a frozen stock vial was used to inoculate a YPLac agar plate (YPLac + agar 2%), which was stored at 4 °C for up to 1 month.

WP with approximately 65 g L^{-1} of lactose was supplied by Carbery (Carbery Food Ingredients, Ballineen, Co. Cork, Ireland) in 3L containers. It was stored at -20 °C freezer upon arrival. When needed, WP was slowly defrosted. Unless stated, WP was first autoclaved in 1 L Duran bottles (15 min, 121 °C). The resulting autoclaved liquid was decanted and filtered through a 0.45 µm membrane by vacuum filtration in aseptic conditions and stored up to two weeks at 4 °C before being used.

Minimal Medium (MM) was prepared as previously described by our group (Donzella et al., 2021). Composition of the vitamin and trace element was used as previously reported (Verduyn et al., 1992). MM was supplemented with ethanol 96% v/v until reaching the desired concentration (up to 8% v/v). To assess the ethanol-consumption capacity, ethanol 2% v/v was used as the sole carbon source. To evaluate the ethanol tolerance, lactose 20 g L^{-1} was supplemented as carbon source.

Inoculum was prepared by transferring a single colony from a YPLac plate to 20 mL of autoclaved and decanted WP held in a 50 mL Corning tube. After 18 to 20 hours of growth on an orbital shaker 200 rpm at 30 °C, cells were directly inoculated to the desired initial concentration in the necessary growth medium.

Adaptive laboratory evolution strategy

Four parallel *K. marxianus* cultures (A, B, C, D) were grown in 250 mL flasks containing 100 mL WP supplemented with different concentrations of ethanol (from 1% to 6%

v/v), starting at A_{600} = 0.01. Cultures were grown for an average of 8 generations (A_{600} 3 to 4) and reinoculated into fresh WP supplemented with ethanol. The number of generations was calculated according to Equation 2.

$$generations = \log_2\left(\frac{A_{final}}{A_{initial}}\right)$$

Equation 2: Calculation of the number of generations grown during the ALE experiment. $A_{initial}$, A_{600} at inoculation; A_{final}, final A_{600} before transfer.

The process began with WP supplemented with 1% v/v ethanol. The content of ethanol in the WP was sequentially increased by 1% as indicated in Table S1 (e.g. to 2% v/v after 59 generations and 6 flask transfers). When growing on WP supplemented with 4% v/v ethanol or higher, cultures were cycled back to WP supplemented with 1% v/v ethanol for 8 generations (1 overnight culture) prior to increasing the concentration to minimise potential fitness trade-offs. This stepwise process continued until WP supplemented with a final concentration of 6% v/v ethanol was reached and maintained for five more cycles, concluding at approximately 600 generations; detail of daily measurements available in Table S1. To characterise the evolved populations, 1 mL from each flask was serially diluted and plated on WP 5% ethanol agar. From each plate, 10 random colonies (40 in total) were picked, and growth was assessed in liquid WP supplemented with 5% ethanol v/v. The 3 best growing strains were selected for a final growth assay (96 well plate format) where they were assessed on WP, WP with 5% ethanol v/v and WP with 6% ethanol v/v and their ethanol production was also evaluated. The best growing and producing isolate overall from each flask (EvA, EvB, EvC, EvD) was selected for small-scale fermentations.

Growth assays

To prepare spot assays, first yeast strains were grown overnight by inoculating one colony from a previously grown YPLac plate into 5 mL of YPLac medium. Once grown, cultures were normalised by measuring A_{600} and collecting an aliquot representing 1 unit of A_{600}. Cells were washed with sterile water and serially diluted. To prepare the spot assay, aliquots of 5 µL of each dilution were placed on the plate and left to dry

under biosafety cabinet. Once dried (~15 minutes), plates were incubated at 30 °C for 24 hours.

For flask fermentation experiments, 250 mL flasks were filled with 150 mL of autoclaved and filtered WP. When needed, ethanol was supplemented 6% v/v. Flasks were inoculated at A_{600} 0.1 and kept for 48 hours on an orbital shaker at 200 rpm and 30 °C. Growth was followed by recording A_{600} in order to compare evolved strains and WT.

Re-pitching experiments were performed in an Alcohol Fermentation Monitor (AFM, Applikon, Delft, Netherlands). cultivations were performed in triplicate mimicking the re-pitched batch performed at Carbery site. Raw WP (not autoclaved) was used as substrate and Duran bottles were filled to 400 mL. Agitation was kept at 50 rpm, temperature maintained at 34 °C, pH was left uncontrolled. The CO_2 is stoichiometrically coupled with carbon-source conversion. Re-pitching was performed when CO_2 production rate plummeted (around 12 hours of fermentation) by centrifuging the fermented WP (5000 rpm 5 minutes) and resuspending the biomass in fresh WP. Samples for ethanol quantification were taken at the end of each cycle. CDW was measured by sampling 10 mL from each bottle at the end of each cycle, spinning down 10 mL in a pre-weighed 50 mL Corning tube, removing supernatant and keeping the tube in a 60 °C incubator until biomass was completely dry.

Bioreactor cultivations were performed in batch by triplicate using an EZcontroller bioreactor (Applikon, Delft, Netherlands). The vessel capacity is 4 L, and each tank was filled up to 3.5 L with WP, supplemented with 0.1 mL L^{-1} of sterile Antifoam C Emulsion (1003171934, Sigma-Aldrich, USA). Agitation was kept at 300 rpm, temperature maintained at 34 °C, pH was left uncontrolled, air was introduced with an external air pump by applying a total flow of 0.5 L min^{-1}. Growth was followed by CDW measurement.

Industrial plant trials

Industrial plant trials were conducted at the Carbery alcohol plant using 8–10 fermentation vessels, each containing approximately 120 000 L of raw WP. The

process was operated according to the standard Carbery process previously described (Ling, 2008), employing *K. marxianus* EvD. The WP was concentrated by RO to increase the lactose content to desired level for fermentation (specific gravity ~1.024 kg L^{-1}).

K. marxianus EvD was propagated in WP supplemented with 0.1% yeast extract (Carbery) under aerobic conditions in dedicated propagation vessels and used to inoculate the fermentation vessel. Fermentations were performed in multiple batch vessels at 34 °C with low-rate air sparging and lasted 12–20 hours. WP specific gravity decreased from approximately 1.024 to 1.008 kg L^{-1} during fermentation, reflecting the conversion of lactose to ethanol.

At the end of each cycle, yeast biomass was recovered by decantation and re-pitched into subsequent batches to maintain high cell density and productivity. Re-pitching was discontinued when specific gravity plateaued and/or pH dropped below 4.5. Following yeast removal, the fermented WP was stored prior to ethanol recovery by distillation.

Analytical methods

Concentration of lactose and ethanol was determined by HPLC. An Agilent 1200 HPLC system (Agilent Technologies, CA, USA) equipped with a refractive index detector and a REZEX 8 µL 8% H+ organic acid column (300 × 7.8 mm) (Phenomenex, CA, USA) were used to separate and detect the analytes. The column was eluted with 0.01 N H_2SO_4 at a flow rate of 0.6 mL min^{-1}. The temperature of the column was maintained at 65 °C.

Statistical analyses

Statistical analyses were performed in Python 3.11 using the SciPy and Pandas libraries. Growth curves from flask experiments were fitted using a logistic growth model (Equation 3).

$$A_{600}(t) = \frac{A_{max}}{1 + e^{-\mu_{max}\,(t - lag)}}$$

Equation 3: Logistic growth model used to fit A_{600} measurements over time, where A_{max} represents the maximum absorbance at 600 nm and μ_{max} the maximum specific growth rate.

Model parameters were estimated for each biological replicate by non-linear least squares fitting using the curve fit function from the SciPy library. The fitted parameter μ was taken as the maximum specific growth rate (μ_{max}), while the lag parameter corresponds to the time of maximal growth acceleration (inflection point), used here as an estimate of lag phase duration. Levene's test indicated unequal variances between groups. Differences between strains (lag and μ_{max} values) were assessed using Welch's t-test, which is robust to unequal variances and small sample sizes, applying a statistical significance of $p < 0.05$.

Yield data were analysed as follows: basic descriptive statistics (mean, standard deviation, and coefficient of variation, CV) were calculated. Normality of each group was assessed using the Shapiro–Wilk test, and homogeneity of variances was evaluated using Levene's test (median-centred).

Depending on these assumptions, differences between groups were tested using either Student's t-test (equal variances) or Welch's t-test (unequal variances). Statistical significance was defined at $p < 0.05$.

Robustness during industrial trials was assessed using the coefficient of variation (CV), with 95% confidence intervals estimated by bootstrap resampling (10,000 iterations). Differences in overall distribution between groups were additionally evaluated using the Kolmogorov–Smirnov test. Statistical significance was defined at $p < 0.05$.

Results

To obtain an ethanol-tolerant *K. marxianus* strain, an ALE strategy, as described in the Methods section, was implemented (Fig. 2a). Four independent cultures of *K. marxianus* (A, B, C, D) were initially grown in WP supplemented with 1% ethanol v/v and gradually adapted to a final concentration of WP supplemented with 6% ethanol v/v. After approximately 600 generations, individual isolates from each population were screened for ethanol tolerance (Fig. 2b). This strategy was adopted since ethanol is a common stressor that is known to activate different response pathways. By selecting using ethanol, the possibility of obtaining more robust strains is increased. The isolates exhibiting the highest growth and ethanol production from each flask (called EvA, EvB, EvC, and EvD) were selected for subsequent small-scale fermentation experiments.

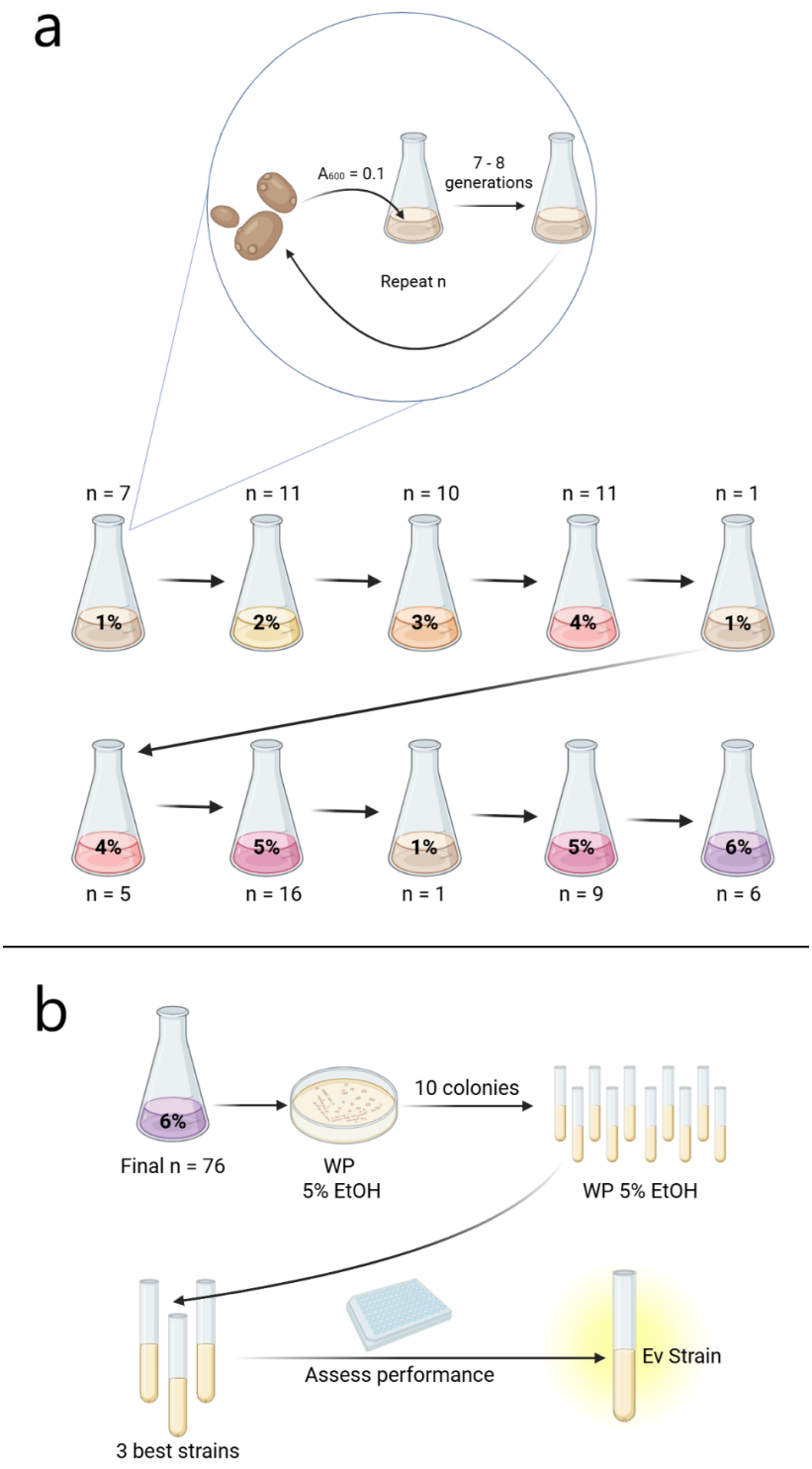


Fig. 2. ALE strategy. a) Sequential ALE strategy adopted to obtain an ethanol tolerant strain. WT *K. marxianus* inoculated in fresh WP supplemented with ethanol (1% to 6% v/v). Once 7 to 8 generations were reached, a transfer to fresh WP with ethanol was done. This was repeated n times. After n times, the next stage with increased ethanol % was started until reaching 6% ethanol v/v. To avoid trade-off effects, during selection at 4% and 5% ethanol v/v a cycle back to 1% was introduced. **b)** Selection of the evolved (Ev) strains at the end of ALE.

To assess the ethanol tolerance of these evolved yeast strains (Fig. 2), we grew them on WP supplemented with ethanol 6% v/v (WP 6%) and compared them to the WT (Fig. 3a). Based on flask growth, all the evolved strains grew faster in WP 6% compared to the WT. In particular, EvD showed the best performance with reduced lag-phase and faster growth. To quantitatively assess these differences, growth curves were fitted to a logistic model and the parameters μ_{max} and lag phase duration (Lag) were extracted for each strain (Fig. 3b, Table S2). For each fitted curve, μ_{max} corresponds directly to the fitted growth rate parameter (μ) of the logistic model, while Lag corresponds to the fitted inflection point parameter (lag), representing the transition from lag phase to exponential growth. Welch's t-test was applied showing that, while the differences in lag phase were not statistically significant ($p > 0.05$) the μ_{max} of EvD was significantly higher than that of the WT ($p = 0.009$), showing that the evolved strain EvD displays an improved growth rate under ethanol stress. We also checked that growth rate in WP without ethanol was not compromised in the evolved strains compared to the WT. Growth in WP with no ethanol (Fig. 3c) showed that all the strains behaved similarly, with no observable differences in lag-phase or growth rate. Finally, to confirm that the increased ethanol tolerance was not due to ethanol, we compared growth on ethanol as the sole carbon source (Fig. 3d), since enhanced ethanol tolerance could otherwise reflect an increased ability to degrade ethanol. Spot assays on MM with 2% ethanol (v/v) showed no observable difference between EvA, EvC and EvD strains and the WT. Taking into account biomass yield, growth rate and lag phase duration, EvD strain was considered the most tolerant and selected for further studies.

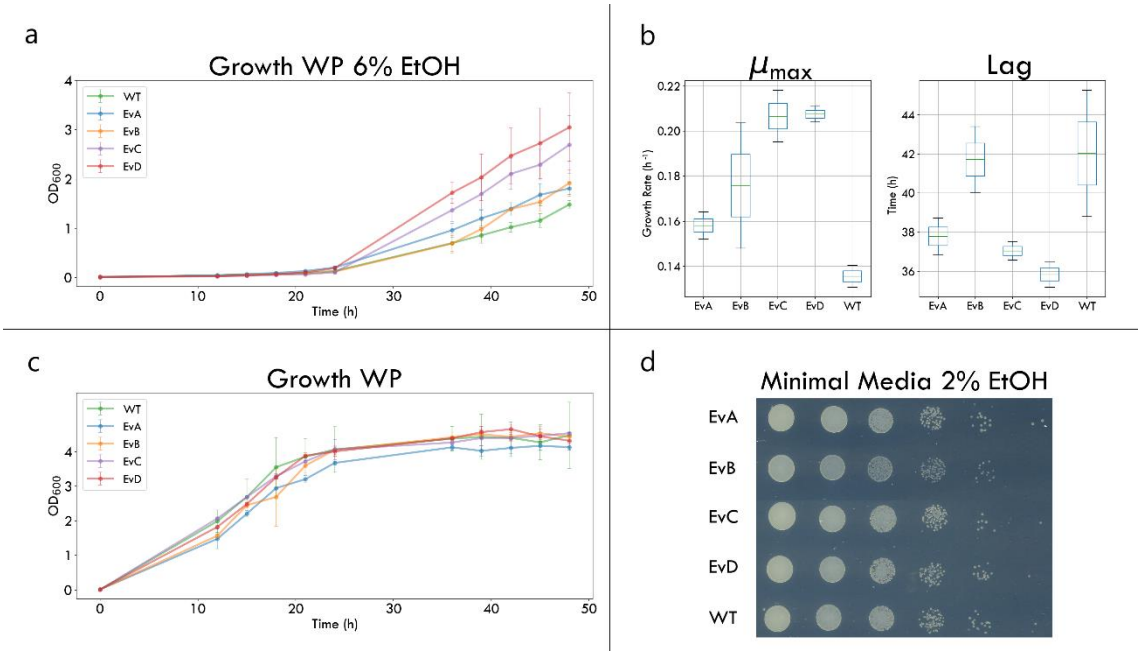


Fig. 3. Characterisation of evolved strains. a) Comparison of the growth of *K. marxianus* WT to 4 ALE isolates (EvA, EvB, EvC and EvD) on WP supplemented with 6% ethanol v/v. **b)** Boxplots of growth parameters extracted from logistic model fitting: μ_{max} and lag phase duration (Lag). EvC and EvD displayed a significant higher μ_{max} and compared to WT. **c)** Growth curves in WP without ethanol, showing no trade-off in baseline growth among evolved strains compared to WT. **d)** Ability of WT and the isolates EvA, EvB, EvC and EvD to grow on MM with ethanol 2% v/v as the sole carbon source, showing no difference in ethanol utilisation ability between WT and evolved strains.

With the purpose of bringing the EvD strain to the industrial set up at Carbery, further scale-up studies were performed in order to assess whether there were any trade-off effects, and to confirm suitability. We first compared the growth of both strains growing on WP in 4L bioreactors (Fig. 4a) and we observed no significant difference between the two strains in terms of growth rate as well as yield (Fig. 4b), with the WT and EvD strains showing similar average yields of 0.424 and 0.42 g of ethanol per g^{-1} of lactose respectively, confirming that the evolved strain is indeed capable of fermenting lactose as efficiently as the parental strain in larger scales. We then attempted to mimic Carbery's conditions and re-pitching approach in an AFM and ran both strains EvD and WT in parallel. We observed that a cyclic growth based on re-pitching was possible using this approach (Fig. 4c), with stable performance along the 6 cycles performed with both strains. We compared the ethanol yield across the

different cycles and observed that EvD performs as well as the WT in this assay, achieving respectively yields of 0.437 and 0.431 ethanol g lactose g^{-1} (Fig. 4d).

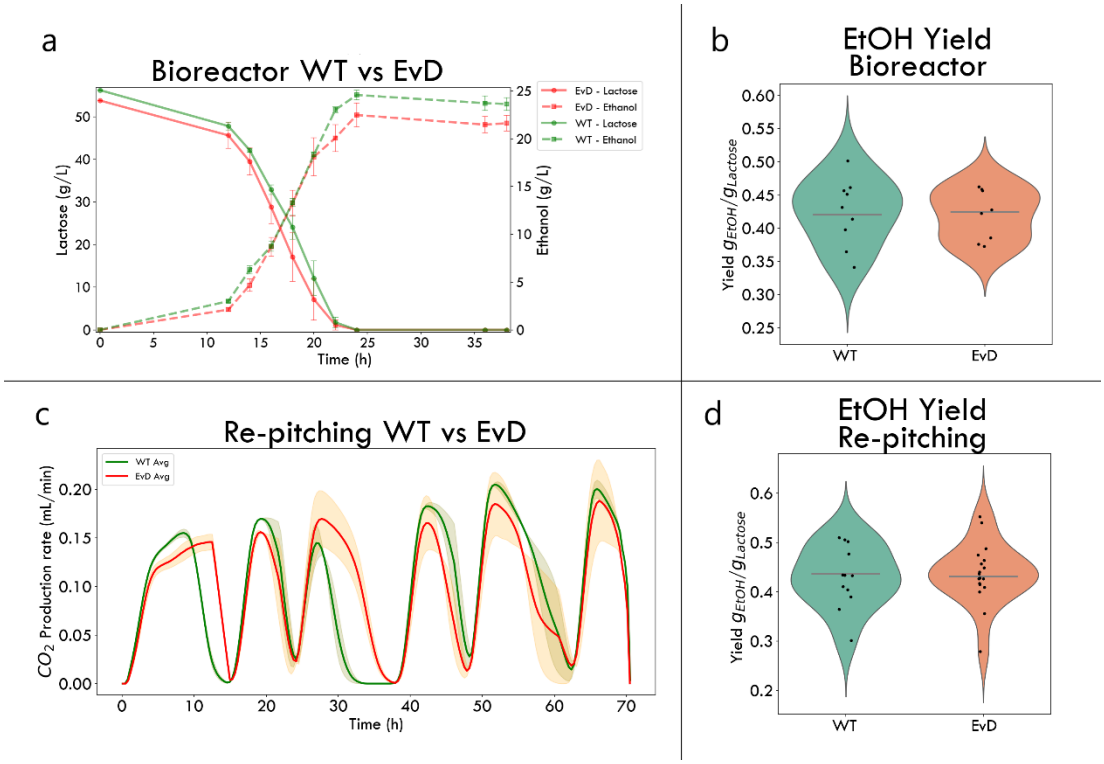


Fig. 4. Growth of WT and EvD in Bioreactor and re-pitched bottles. a) Evolution of lactose consumption (▪, solid lines) and ethanol production (•, dashed lines,) from top-bench bioreactor growth curves of WT (green) and EvD (red) growing on WP. **b)** Violin distribution displaying calculated ethanol yield of each bioreactor for WT strain (green) and EvD (red). Average of each distribution is displayed as a grey line in each violin. **c)** Evolution of the CO_2 production rate in the AFM which serves as a proxy of the fermentation process. To emulate Carbery re-pitching process biomass was separated and re-pitched sequentially up to 6 cycles. WT (green) and EvD (red) performed 6 re-pitching cycles. **d)** Violin distribution displaying calculated ethanol yield for each AFM cycle for WT (green) and EvD (red). Average of each distribution is displayed as a grey line in each violin.

With a strain capable of withstanding higher ethanol stress, a set of trials to compare EvD to the WT was planned at Carbery site following their Standard Operation Procedure (Fig. 1b). Each trial comprised a set of cycles. Each cycle consisted of (i) inoculation of fresh WP, (ii) fermentation until lactose exhaustion, (iii) biomass set down, (iv) supernatant transfer and (v) vessel refill. Cycles were repeated until yeast exhaustion (typically 4 to 5 cycles). Samples to measure lactose and ethanol were taken at start and every 3 hours until the cycle was finished. The concentration of lactose and ethanol allowed us to clearly identify each cycle (Fig. S1). In total, across six independent trials, the WT strain completed 23 cycles while EvD completed 30 cycles. To compare the performance of the strains, the ethanol yield of each cycle

was evaluated (Fig. 5a). Both strains presented similar average yields, with a marginally higher average of EvD than WT (0.324 vs. 0.285 g of ethanol per g^{-1} of lactose, p = 0.12). However, the yield distribution of the EvD strain appeared more concentrated in the upper part of the violin indicating fewer low-yield cycles, which could translate into a more robust strain. To assess this observation quantitatively, the CV was used as a measure of yield stability, revealing lower variability in EvD (CV = 0.238) compared to WT (CV = 0.362). Additionally, the Kolmogorov–Smirnov test confirmed that the overall yield distributions differed significantly between strains (p = 0.015), supporting the conclusion that EvD performs more consistently across cycles. We also observed that the EvD strain is capable of performing more re-pitching cycles than the WT, with an average of 5 cycles per run of the EvD compared to 3.8 cycles of the WT (Fig. 5b). Taken together, these results indicate that EvD not only maintains comparable ethanol yields to the WT but also displays greater process robustness, sustaining productive performance over more successive cycles.

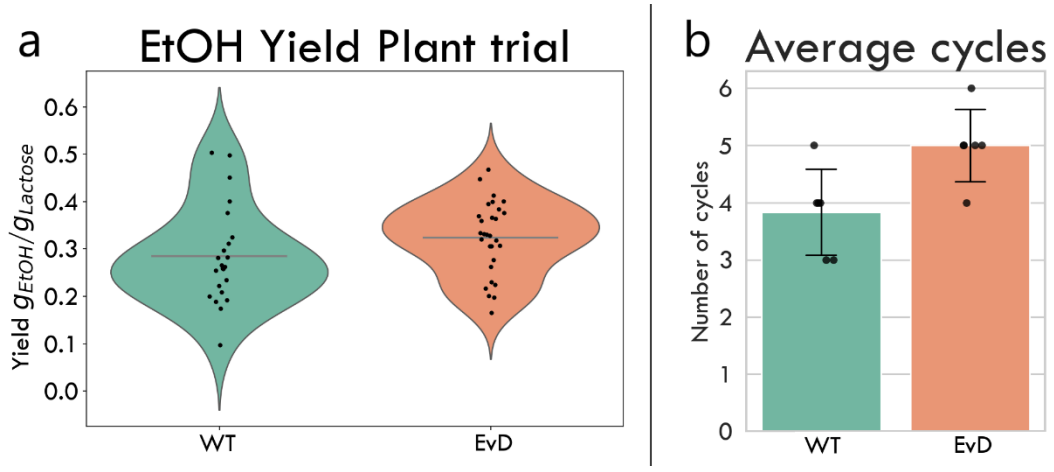


Fig 5. Ethanol yield and average cycle number of WT and EvD during Carbery plant trials. a) Violin distribution displaying calculated ethanol yield for each cycle performed during Carbery plant trials for WT (green) and EvD (red). Average of each distribution is displayed as a grey line in each violin. **b)** Bar graph comparing average number of cycles per run during Carbery plant trials for WT (green) and EvD (red).

Discussion

The dairy yeast *K. marxianus* is a common choice for the production of bioethanol from dairy wastes at industrial scale (Asunis et al., 2020; De Micco et al., 2025; Guimarães et al., 2010). Carbery Food Ingredients has achieved circular manufacturing through a cascade of processes including a fermentation of WP with repeated yeast biomass recovery cycles (re-pitching). In this environment, *K. marxianus* is exposed to fluctuating feedstock properties, accumulation of stressors, and cellular senescence. ALE provides a robust strategy for generating non-GMO strains with enhanced resilience under such industrially relevant stresses (Christer Andersson et al., 2011; da Silveira et al., 2020; Mo et al., 2019; Mullins et al., 2024; Pal & Vij, 2022).

Using ALE (Fig. 2), we generated the EvD strain, which outperformed the parental strain under industrially relevant conditions, displaying a 40% higher μ_{max} in WP supplemented with 6% ethanol (Fig. 3). The molecular mechanisms underlying this improved tolerance remain, however, to be elucidated. Ethanol imposes multifactorial stress on yeast, increasing membrane fluidity and permeability, impairing protein folding, and inducing oxidative stress, which affects growth and viability (de Moura Ferreira et al., 2022). Given this pleiotropic impact, adaptive responses can arise through diverse mutations affecting various physiological pathways. Consistent with this, previous ALE studies in *K. marxianus* have shown that tolerance enhancement often results from regulatory rewiring rather than single-gene mutations. For instance, Mo et al., 2019 obtained an evolved strain (KM-100d) after 100 days in medium supplemented with 6% ethanol v/v, which exhibited broad stress resistance linked to transcriptomic reprogramming and regulatory-region mutations that mimicked an “ethanol-induced” expression profile even in ethanol-free conditions. Similarly, da Silveira et al., 2020 obtained an evolved strain (ETS4) after growing *K. marxianus* in medium supplemented with 4% ethanol v/v for 85 days, identifying mutations in promoter and terminator regions of genes related to lipid metabolism (e.g., *RRI1*, *KLMA_10136*, *PXA2*). These changes were correlated with increased monounsaturated fatty acids and ergosterol, likely contributing to improved membrane stability. In another study, Lertwattanasakul et al., 2023

isolated *K. marxianus* resistant colonies after selecting under heat stress. In the resulting heat-tolerant strains, they identified mutations disrupting *KLMA_50490*, a homolog of *S. cerevisiae SWI5*, suggesting that tolerance evolution can involve cell-cycle or transcriptional regulators. Collectively, these studies suggest that ethanol tolerance in *K. marxianus* typically results from complex regulatory and metabolic adaptations rather than single structural mutations, and we anticipate that similar mechanisms contribute to the phenotype of the EvD strain. However, this remains to be confirmed through genome sequencing and comparative analyses.

Previous ALE studies successfully enhanced ethanol tolerance in *K. marxianus*, but developed strains are usually not tested in industrial setups. It is widely accepted that scale up of laboratory strains is a complex process and it should be considered since the initial development of microbial strains (Garcia-Ochoa & Gomez, 2009; Madhavan et al., 2021; Mayer et al., 2023). While ALE-derived strains are often not tested under industrial conditions, we evaluated the performance of the evolved strain EvD both at laboratory scale, using conditions designed to mimic industrial operations, and in a full-scale industrial plant at Carbery. Ethanol yields of 0.43 and 0.42 g ethanol g^{-1} lactose were obtained in AFM and bioreactors, respectively, corresponding to approximately 80% of the theoretical maximum yield (0.538 g ethanol g^{-1} lactose). These values fall within the upper range of yields reported for *K. marxianus* fermenting dairy substrates (0.35–0.48 g ethanol g^{-1} lactose), which are known to depend on aeration, substrate concentration, and nutrient supplementation (Koukoumaki et al., 2024; Ohstrom et al., 2023; Silveira et al., 2005). Although direct comparison with previous studies is complicated by variability in operating conditions, the small-scale results obtained here are consistent with reported flask and re-pitched bioreactor fermentations of dairy substrates by *K. marxianus* (De Micco et al., 2025; Díez-Antolínez et al., 2018; Pal & Vij, 2022; Saini et al., 2017).

In contrast, the full-scale industrial evaluation in Carbery's plant occasionally exhibited lower yields (<0.35 g ethanol g^{-1} lactose), reflecting the intrinsic complexity of large-scale plant operations, including incomplete mixing, heterogeneity in process conditions, and sampling variability. Importantly, these effects impacted both the

evolved and wild-type strains and therefore did not alter the overall comparative interpretation.

Overall, the evolved strain EvD demonstrated enhanced robustness, sustaining more fermentation cycles with greater yield consistency (Fig. 5), which may reduce propagation and processing costs. WP fermentation is often exposed to fluctuations in operational conditions such as variable lactose concentration, nitrogen composition, temperature shifts, or oxygen transfer efficiency, which can strongly affect strain performance. The enhanced tolerance of the EvD strain indicates greater operational reliability under real plant conditions, and its non-GMO status, together with direct derivation from the WT strain, facilitates integration into existing processes. Nevertheless, both genetic characterisation and process refinement will be key to confirming the long-term robustness and scalability of the EvD strain for full industrial implementation.

More broadly, ALE offers a complementary approach to metabolic engineering, enabling the evolution of polygenic adaptive traits under industrially relevant stresses without introducing recombinant DNA, which is advantageous for food applications (Christer Andersson et al., 2011; Dragosits & Mattanovich, 2013; Mavrommati et al., 2022; More et al., 2022; Mullins et al., 2024). Combining ALE with rational engineering could accelerate strain development further, integrating the strengths of empirical selection and targeted design (Jiang et al., 2021; Wang et al., 2017; Zhou et al., 2023). Beyond WP valorisation, similar evolutionary strategies can enhance tolerance and productivity in other stress-intensive bioconversions, including lignocellulosic feedstocks or organic acid production (Gan et al., 2022; Liu et al., 2021; Ren et al., 2024). From a sustainability standpoint, the use of naturally evolved, robust strains capable of valorising renewable, low-cost substrates contributes to waste minimisation, improved resource efficiency, and advancement of circular bioeconomy objectives. Future efforts should focus on elucidating the genetic basis of the evolved phenotype and integrating ALE with targeted engineering to accelerate the development of next-generation, sustainable microbial platforms for bioethanol and beyond.

Conclusion

This study demonstrates that ALE is a highly effective, non-recombinant strategy for enhancing the industrial robustness of *K. marxianus*. By applying long-term selective pressure for ethanol tolerance, we isolated a variant, EvD, that exhibited better growth under ethanol stress conditions without compromising its native metabolic capabilities. Crucially, this enhanced strain successfully translated from the laboratory to Carbery production plant, where EvD demonstrated greater yield stability and an extended re-pitching lifespan compared to the parental strain. These findings validate the use of ALE not only to overcome specific physiological bottlenecks but also to deliver robust, clean-label microbial solutions compatible with the regulatory and operational demands of the food and dairy sectors. Ultimately, this work contributes to the advancement of the circular bioeconomy by improving the efficiency and economic viability of converting dairy side-streams into renewable fuels.

References

Asunis, F., De Gioannis, G., Dessì, P., Isipato, M., Lens, P. N. L., Muntoni, A., Polettini, A., Pomi, R., Rossi, A., & Spiga, D. (2020). The dairy biorefinery: Integrating treatment processes for cheese whey valorisation. *Journal of Environmental Management*, *276*, 111240. https://doi.org/10.1016/J.JENVMAN.2020.111240

Capper, J. L., Cady, R. A., & Bauman, D. E. (2009). The environmental impact of dairy production: 1944 compared with 2007. *Journal of Animal Science*, *87*(6), 2160–2167. https://doi.org/10.2527/JAS.2009-1781

Christer Andersson, H., Arpaia, S., Bartsch, D., Casacuberta, J., Davies, H., du Jardin, P., Flachowsky, G., Herman, L., Jones, H., Kärenlampi, S., Kiss, J., Kleter, G., Kuiper, H., Messéan, A., Magne Nielsen, K., Perry, J., Pöting, A., Sweet, J., Tebbe, C., … Wal, J.-M. (2011). Guidance on the risk assessment of genetically modified microorganisms and their products intended for food and feed use. *EFSA Journal*, *9*(6), 2193. https://doi.org/10.2903/J.EFSA.2011.2193

Costa, D. A., De Souza, C. J. A., Costa, P. S., Rodrigues, M. Q. R. B., Dos Santos, A. F., Lopes, M. R., Genier, H. L. A., Silveira, W. B., & Fietto, L. G. (2014). Physiological characterization of thermotolerant yeast for cellulosic ethanol production. *Applied Microbiology and Biotechnology*, *98*(8), 3829–3840. https://doi.org/10.1007/S00253-014-5580-3

da Silveira, F. A., de Oliveira Soares, D. L., Bang, K. W., Balbino, T. R., de Moura Ferreira, M. A., Diniz, R. H. S., de Lima, L. A., Brandão, M. M., Villas-Bôas, S. G., & da Silveira, W. B. (2020). Assessment of ethanol tolerance of Kluyveromyces marxianus CCT 7735 selected by adaptive laboratory evolution. *Applied Microbiology and Biotechnology*, *104*(17), 7483–7494. https://doi.org/10.1007/s00253-020-10768-9

da Silveira Garcia, L., Fernandes Bettencourt, A., Fernandes Aires, J., Andretta, I., & Fischer, V. (2024). Seasonal stability of raw bovine milk: a systematic review and meta-analysis. *Biological Rhythm Research*, *55*(1), 30–44. https://doi.org/10.1080/09291016.2023.2284545

De Micco, C., Radice, M., Di Fidio, N., Agrimi, G., & Pisano, I. (2025). Physiological characterization of Kluyveromyces marxianus strains for the development of a microbial platform to obtain bioethanol from cheese whey permeate. *Frontiers in Microbiology*, *16*, 1663736. https://doi.org/10.3389/FMICB.2025.1663736

de Moura Ferreira, M. A., da Silveira, F. A., & da Silveira, W. B. (2022). Ethanol stress responses in Kluyveromyces marxianus: current knowledge and perspectives. *Applied Microbiology and Biotechnology*, *106*(4), 1341–1353. https://doi.org/10.1007/S00253-022-11799-0

Díez-Antolínez, R., Hijosa-Valsero, M., Paniagua-García, A. I., Garita-Cambronero, J., & Gómez, X. (2018). Yeast screening and cell immobilization on inert supports for ethanol production from cheese whey permeate with high lactose loads. *PLOS ONE*, *13*(12), e0210002. https://doi.org/10.1371/JOURNAL.PONE.0210002

Donzella, L., Varela, J. A., Sousa, M. J., & Morrissey, J. P. (2021). Identification of novel pentose transporters in *Kluyveromyces marxianus* using a new screening platform. *FEMS Yeast Research*, *21*(4), 26. https://doi.org/10.1093/femsyr/foab026

Dragosits, M., & Mattanovich, D. (2013). Adaptive laboratory evolution - principles and applications for biotechnology. In *Microbial Cell Factories* (Vol. 12, Issue 1, pp. 1–17). BioMed Central. https://doi.org/10.1186/1475-2859-12-64

European Commission. (2019). *The European Green Deal*. https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en

Finnegan, W., & Goggins, J. (2021). Environmental impact of the dairy industry. *Environmental Impact of Agro-Food Industry and Food Consumption*, 129–146. https://doi.org/10.1016/B978-0-12-821363-6.00004-7

Gan, Y., Bai, M., Lin, X., Liu, K., Huang, B., Jiang, X., Liu, Y., & Gao, C. (2022). Improvement of macrolactins production by the genetic adaptation of Bacillus siamensis A72 to saline stress via adaptive laboratory evolution. *Microbial Cell Factories 2022 21:1*, *21*(1), 1–15. https://doi.org/10.1186/S12934-022-01871-9

Garcia-Ochoa, F., & Gomez, E. (2009). Bioreactor scale-up and oxygen transfer rate in microbial processes: An overview. In *Biotechnology Advances* (Vol. 27, Issue 2, pp. 153–176). Elsevier. https://doi.org/10.1016/j.biotechadv.2008.10.006

Guimarães, P. M. R., Teixeira, J. A., & Domingues, L. (2010). Fermentation of lactose to bio-ethanol by yeasts as part of integrated solutions for the valorisation of cheese whey. *Biotechnology Advances*, *28*(3), 375–384. https://doi.org/10.1016/J.BIOTECHADV.2010.02.002

Guzmán-Luna, P., Mauricio-Iglesias, M., Flysjö, A., & Hospido, A. (2022). Analysing the interaction between the dairy sector and climate change from a life cycle perspective: A review. *Trends in Food Science & Technology*, *126*, 168–179. https://doi.org/10.1016/J.TIFS.2021.09.001

Jiang, W., Hernández Villamor, D., Peng, H., Chen, J., Liu, L., Haritos, V., & Ledesma-Amaro, R. (2021). Metabolic engineering strategies to enable microbial utilization of C1 feedstocks. *Nature Chemical Biology 2021 17:8*, *17*(8), 845–855. https://doi.org/10.1038/s41589-021-00836-0

Koukoumaki, D. I., Papanikolaou, S., Ioannou, Z., Mourtzinos, I., & Sarris, D. (2024). Single-Cell Protein and Ethanol Production of a Newly Isolated Kluyveromyces marxianus Strain through Cheese Whey Valorization. *Foods*, *13*(12), 1892. https://doi.org/10.3390/FOODS13121892

Lee, S. R., & Kim, P. (2020). Current Status and Applications of Adaptive Laboratory Evolution in Industrial Microorganisms. *Journal of Microbiology and Biotechnology*, *30*(6), 793. https://doi.org/10.4014/JMB.2003.03072

Lertwattanasakul, N., Pattanakittivorakul, S., Nitiyon, S., Matsutani, M., Oguchi, A., Hirata, K., Kosaka, T., Limtong, S., & Yamada, M. (2023). Mutants with Enhanced Multi-Stress Tolerance of Kluyveromyces marxianus and Their Ability for Ethanol Fermentation. *Fuels 2023, Vol. 4, Pages 469-483*, *4*(4), 469–483. https://doi.org/10.3390/FUELS4040029

Ling, K. C. (2008). *Whey to Ethanol: A Biofuel Role for Dairy Cooperatives?* https://doi.org/10.22004/AG.ECON.280102

Liu, Z., Radi, M., Mohamed, E. T. T., Feist, A. M., Dragone, G., & Mussatto, S. I. (2021). Adaptive laboratory evolution of Rhodosporidium toruloides to inhibitors derived from lignocellulosic biomass and genetic variations behind evolution. *Bioresource Technology*, *333*, 125171. https://doi.org/10.1016/J.BIORTECH.2021.125171

Madhavan, A., Arun, K. B., Sindhu, R., Krishnamoorthy, J., Reshmy, R., Sirohi, R., Pugazhendi, A., Awasthi, M. K., Szakacs, G., & Binod, P. (2021). Customized yeast cell factories for biopharmaceuticals: from cell engineering to process scale up. *Microbial Cell Factories 2021 20:1*, *20*(1), 1–17. https://doi.org/10.1186/S12934-021-01617-Z

Mavrommati, M., Daskalaki, A., Papanikolaou, S., & Aggelis, G. (2022). Adaptive laboratory evolution principles and applications in industrial biotechnology. *Biotechnology Advances*, *54*. https://doi.org/10.1016/j.biotechadv.2021.107795

Mayer, F., Haslinger, B., Shpylovyi, A., Weber, A., Windberger, U., Albrecht, B., Hahn, R., Cserjan-Puschmann, M., & Striedner, G. (2023). Scale-related process heterogeneities change properties of high-cell-density fermentation broths demonstrated with Escherichia coli B and K-12 strains. *Journal of Chemical Technology & Biotechnology*, *98*(6), 1443–1452. https://doi.org/10.1002/JCTB.7363

Mo, W., Wang, M., Zhan, R., Yu, Y., He, Y., & Lu, H. (2019). Kluyveromyces marxianus developing ethanol tolerance during adaptive evolution with significant improvements of multiple pathways. *Biotechnology for Biofuels*, *12*(1), 1–15. https://doi.org/10.1186/s13068-019-1393-z

More, S., Bampidis, V., Benford, D., Bragard, C., Halldorsson, T., Hernández-Jerez, A., Bennekou, S. H., Koutsoumanis, K., Lambré, C., Machera, K., Mullins, E., Nielsen, S. S., Schlatter, J., Schrenk, D., Turck, D., Younes, M., Herman, L., Pelaez, C., van Loveren, H., … Cocconcelli, P. S. (2022). Evaluation of existing guidelines for their adequacy for the food and feed risk assessment of microorganisms obtained

through synthetic biology. *EFSA Journal*, *20*(8), e07479. https://doi.org/10.2903/J.EFSA.2022.7479

Mukherjee, P., Raj, N., & Sivaprakasam, S. (2023). Harnessing valorization potential of whey permeate for D-lactic acid production using lactic acid bacteria. *Biomass Conversion and Biorefinery 2023 13:17*, *13*(17), 15639–15658. https://doi.org/10.1007/S13399-023-05038-3

Mullins, E., Bresson, J. L., Dewhurst, I. C., Epstein, M. M., Firbank, L. G., Guerche, P., Hejatko, J., Moreno, F. J., Naegeli, H., Nogué, F., Rostoks, N., Sánchez Serrano, J. J., Savoini, G., Veromann, E., Veronesi, F., Cocconcelli, P. S., Glandorf, D., Herman, L., Jimenez Saiz, R., … Dalmay, T. (2024). New developments in biotechnology applied to microorganisms. *EFSA Journal*, *22*(7), e8895. https://doi.org/10.2903/J.EFSA.2024.8895

Ohstrom, A. M., Buck, A. E., Du, X., & Wee, J. (2023). Evaluation of Kluyveromyces spp. for conversion of lactose in different types of whey from dairy processing waste into ethanol. *Frontiers in Microbiology*, *14*, 1208284. https://doi.org/10.3389/FMICB.2023.1208284

Ortiz-Merino, R. A., Varela, J. A., Coughlan, A. Y., Hoshida, H., da Silveira, W. B., Wilde, C., Kuijpers, N. G. A., Geertman, J. M., Wolfe, K. H., & Morrissey, J. P. (2018). Ploidy variation in Kluyveromyces marxianus separates dairy and non-dairy isolates. *Frontiers in Genetics*, *9*(MAR), 94. https://doi.org/10.3389/fgene.2018.00094

Pal, U., & Vij, S. (2022). Adaptive evolution of Kluyveromyces marxianus MTCC1389 for high ethanol tolerance. *Biocatalysis and Agricultural Biotechnology*, *45*, 102533. https://doi.org/10.1016/J.BCAB.2022.102533

Pasotti, L., Zucca, S., Casanova, M., Micoli, G., Cusella De Angelis, M. G., & Magni, P. (2017). Fermentation of lactose to ethanol in cheese whey permeate and concentrated permeate by engineered Escherichia coli. *BMC Biotechnology*, *17*(1), 1–12. https://doi.org/10.1186/S12896-017-0369-Y

Pratt, P. L., Bryce, J. H., & Stewart, G. G. (2003). The Effects of Osmotic Pressure and Ethanol on Yeast Viability and Morphology. *Journal of the Institute of Brewing*, *109*(3), 218–228. https://doi.org/10.1002/J.2050-0416.2003.TB00162.X

Ren, J., Zhang, M., Guo, X., Zhou, X., Ding, N., Lei, C., Jia, C., Wang, Y., Zhao, J., Dong, Z., & Lu, D. (2024). Furfural tolerance of mutant Saccharomyces cerevisiae selected via ionizing radiation combined with adaptive laboratory evolution. *Biotechnology for Biofuels and Bioproducts 2024 17:1*, *17*(1), 1–18. https://doi.org/10.1186/S13068-024-02562-W

Saini, P., Beniwal, A., Kokkiligadda, A., & Vij, S. (2017). Evolutionary adaptation of Kluyveromyces marxianus strain for efficient conversion of whey lactose to bioethanol. *Process Biochemistry*, *62*, 69–79. https://doi.org/10.1016/j.procbio.2017.07.013

Silveira, W. B., Passos, F. J. V., Mantovani, H. C., & Passos, F. M. L. (2005). Ethanol production from cheese whey permeate by Kluyveromyces marxianus UFV-3: A flux analysis of oxido-reductive metabolism as a function of lactose concentration and oxygen levels. *Enzyme and Microbial Technology*, *36*(7), 930–936. https://doi.org/10.1016/j.enzmictec.2005.01.018

Tesfaw, A. (2023). The current trends of bioethanol production from cheese whey using yeasts: biological and economical perspectives. *Frontiers in Energy Research*, *11*, 1183035. https://doi.org/10.3389/FENRG.2023.1183035

Varela, J. A., Puricelli, M., Ortiz-Merino, R. A., Giacomobono, R., Braun-Galleani, S., Wolfe, K. H., & Morrissey, J. P. (2019). Origin of Lactose Fermentation in Kluyveromyces lactis by Interspecies Transfer of a Neo-functionalized Gene Cluster during Domestication. *Current Biology*, *29*(24), 4284-4290.e2. https://doi.org/10.1016/j.cub.2019.10.044

Verduyn, C., Postma, E., Scheffers, W. A., & Van Dijken, J. P. (1992). Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast*, *8*(7), 501–517. https://doi.org/10.1002/YEA.320080703

Wang, J., Butler, R. R., Wu, F., Pombert, J. F., Kilbane, J. J., & Stark, B. C. (2017). Enhancement of Microbial Biodesulfurization via Genetic Engineering and Adaptive Evolution. *PLOS ONE*, *12*(1), e0168833. https://doi.org/10.1371/JOURNAL.PONE.0168833

Wauters, R., Britton, S. J., & Verstrepen, K. J. (2021). Old yeasts, young beer—The industrial relevance of yeast chronological life span. *Yeast*, *38*(6), 339–351. https://doi.org/10.1002/YEA.3650

Yaakoub, H., Sanchez, N. S., Ongay-Larios, L., Courdavault, V., Calenda, A., Bouchara, J. P., Coria, R., & Papon, N. (2022). The high osmolarity glycerol (HOG) pathway in fungi†. *Critical Reviews in Microbiology*, *48*(6), 657–695. https://doi.org/10.1080/1040841X.2021.2011834

Zhou, K., Yu, C., Liang, N., Xiao, W., Wang, Y., Yao, M., & Yuan, Y. (2023). Adaptive Evolution and Metabolic Engineering Boost Lycopene Production in Saccharomyces cerevisiae via Enhanced Precursors Supply and Utilization. *Journal of Agricultural and Food Chemistry*, *71*(8), 3821–3831. https://doi.org/10.1021/ACS.JAFC.2C08579

Zotta, T., Solieri, L., Iacumin, L., Picozzi, C., & Gullo, M. (2020). Valorization of cheese whey using microbial fermentations. *Applied Microbiology and Biotechnology 2020 104:7*, *104*(7), 2749–2764. https://doi.org/10.1007/S00253-020-10408-2

Supplementary files

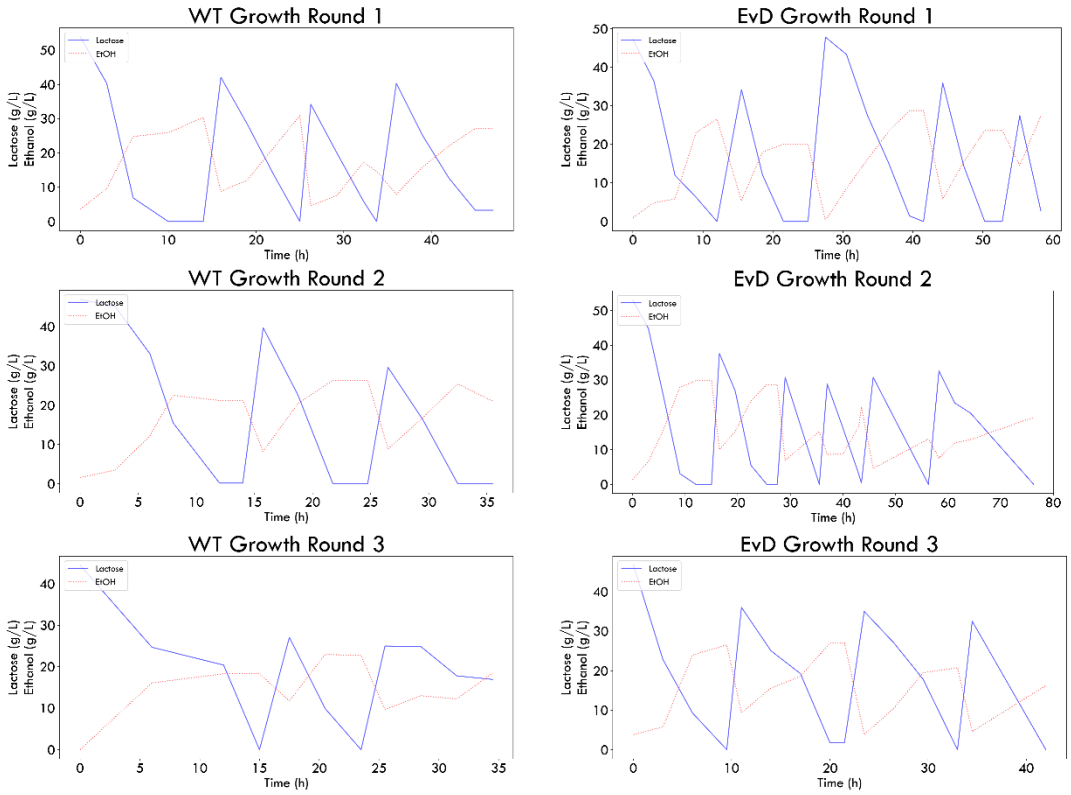


Fig. S1. Growth curves during the plant trials. Example of different trials performed at Carbery site. Lactose (blue) and ethanol (red) follow a cyclic pattern during re-pitching. Each peak in blue represents one cycle.

Flask n	% EtOH	A_{600} A	generations A	A_{600} B	generation B	A_{600} C	generations C	A_{600} D	generations D
0	1%	3.40	8.409390936	3.76	8.554588852	4.2	8.714245518	4.63	8.854868383
1	1%	3.50	8.451211112	5	8.965784285	4.06	8.665335917	4.9	8.936637939
2	1%	3.88	8.599912842	3.82	8.577428828	4.37	8.77148947	3.78	8.562242424
3	1%	1.80	7.491853096	2.08	7.700439718	1.75	7.451211112	2.08	7.700439718
4	1%	5.54	9.113742166	4.08	8.672425342	3.77	8.558420713	3.94	8.622051819
5	1%	4.93	8.945443836	4.89	8.933690655	4.31	8.751544059	3.89	8.603626345
6	1%	3.25	8.344295908	3.55	8.471675214	3.37	8.396604781	3.14	8.294620749
7	2%	1.29	7.010108453	1.644	7.361066489	1.45	7.17990909	1.49	7.21916852
8	2%	3.56	8.475733431	3.36	8.392317423	3.75	8.550746785	3.78	8.562242424
9	2%	3.66	8.515699838	3.28	8.357552005	3.33	8.379378367	3.23	8.335390355
10	2%	3.44	8.426264755	4.01	8.647458426	3.84	8.584962501	3.4	8.409390936
11	2%	1.70	7.409390936	1.38	7.108524457	1.81	7.499845887	1.27	6.988684687
12	2%	4.12	8.686500527	4.83	8.915879379	4.51	8.816983623	4.13	8.689997971
13	2%	3.46	8.434628228	3.98	8.636624621	4.15	8.696967526	3.97	8.632995197
14	2%	2.91	8.184875343	3.79	8.566054038	3.42	8.417852515	5.1	8.994353437
15	2%	1.32	7.044394119	1.2	6.906890596	1.5	7.22881869	1.24	6.95419631
16	2%	3.81	8.573647187	4.13	8.689997971	3.95	8.625708843	3.59	8.487840034
17	2%	0.29	4.832890014	2.98	8.21916852	3.9	8.607330314	3.66	8.515699838
18	3%	3.78	8.562242424	3.4	8.409390936	3.47	8.438791853	3.28	8.357552005
19	3%	0.55	5.781359714	1.1	6.781359714	0.95	6.569855608	1	6.64385619
20	3%	4.66	8.864186145	4.59	8.842350343	4.1	8.6794801	3.87	8.596189756
21	3%	3.39	8.405141463	3.53	8.463524373	3.7	8.531381461	3.48	8.442943496
22	3%	4.05	8.661778098	4.19	8.710806434	4.12	8.686500527	4.23	8.724513853
23	3%	0.63	5.977279923	0.84	6.392317423	0.83	6.375039431	0.81	6.339850003
24	3%	4.04	8.658211483	4.38	8.77478706	4.88	8.930737338	3.92	8.614709844
25	3%	3.75	8.550746785	3.84	8.584962501	4.22	8.721099189	3.64	8.50779464
26	3%	3.41	8.413627929	3.34	8.383704292	3.47	8.438791853	3.81	8.573647187
27	3%	1.23	6.942514505	0.985	6.622051819	0.995	6.636624621	1.05	6.714245518
28	4%	3.25	8.344295908	3.44	8.426264755	3.22	8.330916878	3.44	8.426264755
29	4%	2.56	8	2.56	8	2.56	8	2.47	7.948367232
30	4%	2.95	8.204571144	2.76	8.108524457	2.79	8.124121312	2.95	8.204571144
31	4%	0.29	4.832890014	0.275	4.781359714	0.25	4.64385619	0.675	6.076815597
32	4%	3.80	8.569855608	3.52	8.459431619	3.58	8.483815777	3.64	8.50779464
33	4%	2.64	8.044394119	2.8	8.129283017	2.48	7.95419631	2.2	7.781359714
34	4%	2.27	7.826548487	2.28	7.832890014	2.2	7.781359714	2.6	8.022367813
35	4%	0.50	5.64385619	0.585	5.87036472	0.54	5.754887502	0.48	5.584962501
36	4%	3.21	8.326429487	2.86	8.159871337	3	8.22881869	2.7	8.076815597
37	4%	2.92	8.189824559	2.5	7.965784285	2.55	7.994353437	2.53	7.982993575
38	4%	2.86	8.159871337	2.83	8.144658243	3.37	8.396604781	2.36	7.882643049
39	1%	1.75	7.451211112	1.68	7.392317423	1.77	7.46760555	1.63	7.348728154
40	4%	2.97	8.214319121	2.59	8.016808288	2.71	8.082149041	2.85	8.154818109
41	4%	2.40	7.906890596	2.56	8	2.34	7.87036472	2.36	7.882643049
42	4%	2.02	7.658211483	1.63	7.348728154	1.68	7.392317423	1.62	7.339850003
43	4%	0.32	5	0.32	5	0.595	5.894817763	0.545	5.768184325
44	4%	2.93	8.194756854	2.47	7.948367232	2.85	8.154818109	2.8	8.129283017
45	5%	2.07	7.693486957	1.9	7.569855608	2.18	7.768184325	2.1	7.714245518
46	5%	2.33	7.861086906	2.642	8.045486656	2.334	7.866660751	2.531	7.983563697
47	5%	1.77	7.46760555	2.163	7.756889855	2.181	7.76884596	2.231	7.801546703
48	5%	1.82	7.50779464	1.74	7.442943496	1.75	7.451211112	1.76	7.459431619
49	5%	2.15	7.74819285	1.05	6.714245518	0.94	6.554588852	2.33	7.864186145
50	5%	1.85	7.531381461	1.675	7.388017285	1.7	7.409390936	1.55	7.276124405
51	5%	1.65	7.366322214	1.525	7.252665432	1.615	7.335390355	1.58	7.303780748
52	5%	1.70	7.409390936	1.48	7.209453366	1.9	7.569855608	1.8	7.491853096
53	5%	2.11	7.721099189	2.3	7.845490051	1.6	7.321928095	1.93	7.592457037
54	5%	2.20	7.781359714	2.12	7.727920455	2.13	7.73470962	2.4	7.906890596
55	5%	2.11	7.721099189	2.32	7.857980995	2.78	8.118941073	2.2	7.781359714
56	5%	2.12	7.727920455	2.04	7.672425342	2.22	7.794415866	2.17	7.761551232
57	5%	1.83	7.515699838	1.96	7.614709844	1.58	7.303780748	1.8	7.491853096
58	5%	2.34	7.87036472	2.77	8.113742166	3.07	8.262094845	1.91	7.577428828
59	5%	2.24	7.807354922	2.12	7.727920455	2.22	7.794415866	1.98	7.62935662
60	5%	2.41	7.912889336	2.88	8.169925001	3	8.22881869	2.11	7.721099189

61	1%	4.14	8.693486957	4.77	8.897845456	4.45	8.797661526	4.39	8.77807713
62	5%	2.06	7.686500527	2.09	7.707359132	2.54	7.988684687	2.05	7.6794801
63	5%	2.22	7.794415866	2.15	7.74819285	2.12	7.727920455	2.1	7.714245518
64	5%	3.26	8.348728154	2.4	7.906890596	2.51	7.971543554	2.14	7.741466986
65	5%	2.18	7.768184325	2.12	7.727920455	2.27	7.826548487	2.05	7.6794801
66	5%	2.76	8.108524457	2.61	8.027905997	2.54	7.988684687	2.42	7.918863237
67	5%	2.21	7.787902559	2	7.64385619	1.97	7.622051819	1.94	7.599912842
68	5%	2.07	7.693486957	2.08	7.700439718	2.07	7.693486957	2.13	7.73470962
69	5%	2.14	7.741466986	2.14	7.741466986	2.22	7.794415866	2.13	7.73470962
70	5%	2.52	7.977279923	2.17	7.761551232	2.63	8.038918989	2.94	8.199672345
71	6%	1.93	7.592457037	2.08	7.700439718	4.33	8.758223215	2.46	7.942514505
72	6%	2.11	7.721099189	2.09	7.707359132	3.04	8.247927513	2.96	8.209453366
73	6%	4.11	8.682994584	2.19	7.77478706	3.08	8.266786541	2.65	8.049848549
74	6%	1.18	6.882643049	1	6.64385619	1.6	7.321928095	0.74	6.209453366
75	6%	2.14	7.741466986	2.75	8.103287808	2.95	8.204571144	2.08	7.700439718
76	6%	2.26	7.820178962	2.18	7.768184325	2.28	7.832890014	2.3	7.845490051
		Total A	**601.0188626**	**Total B**	**603.8984893**	**Total C**	**608.8261371**	**Total D**	**605.0927571**

Table S1. Recordings of A_{600} in the 4 flasks (A, B, C, D) used during ALE. The OD was recorder right before each passage.

Script S1. Python script used for analysis of growth performance. This script analyses microbial growth curves by fitting a logistic growth model to experimental data for multiple strains and biological replicates. Model parameters, including maximum optical density (A_{600}), maximum specific growth rate (μ_{max}), and lag phase duration, are estimated for each replicate using non-linear least squares fitting. Extracted parameters are compiled into a summary table for downstream analysis. Statistical differences in growth rate between strains are assessed using Levene's test to evaluate variance homogeneity, followed by Welch's t-test for pairwise comparison. Results are visualised as boxplots of μ_{max} and lag phase across strains, allowing comparison of growth dynamics between conditions.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit
from scipy import stats

# ---------- Growth model (Logistic) ----------
def logistic(t, A, mu, lag):
    return A / (1 + np.exp(-mu * (t - lag)))

# ---------- Load data ----------
file = r"file_path/file.xlsx"
df = pd.read_excel(file)
plt.rcParams['font.family'] = 'Tw Cen MT'

time = df.iloc[:,0].values
strain_names = ["WT", "EvA", "EvB", "EvC", "EvD"]
n_replicates = 2

# ---------- Fit curves & extract parameters ----------
results = []

for i, strain in enumerate(strain_names):
    for r in range(n_replicates):
        ydata = df.iloc[:, 1 + i*n_replicates + r].values
```

```
        # initial guesses
        A0 = max(ydata)
        mu0 = 0.3
        lag0 = time[np.argmax(np.diff(ydata))]

        try:
            popt, _ = curve_fit(logistic, time, ydata, p0=[A0, mu0, lag0],
maxfev=5000)
            A, mu, lag = popt
            results.append({"Strain": strain, "Replicate": r+1,
                            "maxOD": A, "mu_max": mu, "lag": lag})
        except RuntimeError:
            print(f"Fit failed for {strain} replicate {r+1}")

res_df = pd.DataFrame(results)
print(res_df)

# ---------- Statistical analysis ----------
data = res_df[['Strain', 'mu_max']]

# --- 2. Check homogeneity of variances (Levene's test) ---
grouped = [subset['mu_max'].values for _, subset in data.groupby('Strain')]
levene_stat, levene_p = stats.levene(*grouped)
equal_var = levene_p > 0.05
print(f"\n=== VARIANCE EQUALITY (Levene's test) ===\nW={levene_stat:.3f},
p={levene_p:.4f}")

# For EvD vs WT
wt = res_df.loc[res_df['Strain'] == 'WT', 'mu_max']
evd = res_df.loc[res_df['Strain'] == 'EvD', 'mu_max']

# Welch's t-test (unequal variances)
```

```
t_stat, p_val = stats.ttest_ind(evd, wt, equal_var=False)

print(f"Welch’s t-test: t={t_stat:.3f}, p={p_val:.4f}")

# ---------- Plots ----------
# Growth parameters boxplots
fig, axes = plt.subplots(1, 2, figsize=(15,5))

# Map parameter names to descriptive labels with units
y_labels = {
    "mu_max": "Growth Rate (h-¹)",
    "lag": "Lag Time (h)"
}

for i, param in enumerate(["mu_max", "lag"]):
    res_df.boxplot(column=param, by="Strain", ax=axes[i], fontsize=50)
    axes[i].set_title(param, fontsize=40)

    # Set a specific label for each plot
    axes[i].set_ylabel(r"$\mu_{max}$ (h$^{-1}$)", fontsize=50)

plt.suptitle("", fontsize=20)
plt.tight_layout(w_pad=5.0)
plt.show()
```

Script S2. Python script used for visualisation of growth performance from AFM. This script processes and visualises CO_2 production rate data from multiple biological replicates. Time-series data for each replicate are first cleaned and interpolated onto a common time grid to enable direct comparison. Mean profiles and standard deviations are then calculated across replicates for each strain. The results are visualised as averaged curves with shaded regions representing variability, allowing comparison of CO_2 production as proxy of growth.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

# === Load Excel file ===
file_path = r"file_path/file.xlsx"
df = pd.read_excel(file_path)
plt.rcParams['font.family'] = 'Tw Cen MT'

# --- Wild type replicates ---
rep1_wt = df[["h1", "wt1"]].dropna()
rep2_wt = df[["h2", "wt2"]].dropna()
rep3_wt = df[["h3", "wt3"]].dropna()
rep4_wt = df[["h4", "wt4"]].dropna()

time1, co2_1 = rep1_wt["h1"].values, rep1_wt["wt1"].values
time2, co2_2 = rep2_wt["h2"].values, rep2_wt["wt2"].values
time3, co2_3 = rep3_wt["h3"].values, rep3_wt["wt3"].values
time4, co2_4 = rep4_wt["h4"].values, rep4_wt["wt4"].values

# Interpolation grid (common timeline)
common_time = np.linspace(
    max(time1.min(), time2.min(), time3.min(), time4.min()),
    min(time1.max(), time2.max(), time3.max(), time4.max()),
    200
)
interp1 = np.interp(common_time, time1, co2_1)
```

```
interp2 = np.interp(common_time, time2, co2_2)
interp3 = np.interp(common_time, time3, co2_3)
interp4 = np.interp(common_time, time4, co2_4)
avg_wt = (interp1 + interp2 + interp3 + interp4) / 4
wt_std = np.std([interp1, interp2, interp3, interp4], axis=0)

# --- Mutant replicates ---
rep5_mut = df[["h5", "d5"]].dropna()
rep6_mut = df[["h6", "d6"]].dropna()
rep7_mut = df[["h7", "d7"]].dropna()
rep8_mut = df[["h8", "d8"]].dropna()

time5, co2_5 = rep5_mut["h5"].values, rep5_mut["d5"].values
time6, co2_6 = rep6_mut["h6"].values, rep6_mut["d6"].values
time7, co2_7 = rep7_mut["h7"].values, rep7_mut["d7"].values
time8, co2_8 = rep8_mut["h8"].values, rep8_mut["d8"].values

# Interpolation grid for mutant
common_time2 = np.linspace(
    max(time5.min(), time6.min(), time7.min(), time8.min()),
    min(time5.max(), time6.max(), time7.max(), time8.max()),
    200
)
interp5 = np.interp(common_time2, time5, co2_5)
interp6 = np.interp(common_time2, time6, co2_6)
interp7 = np.interp(common_time2, time7, co2_7)
interp8 = np.interp(common_time2, time8, co2_8)
avg_mut = (interp5 + interp6 + interp7 + interp8) / 4
mut_std = np.std([interp5, interp6, interp7, interp8], axis=0)

# === Plot ===
plt.figure(figsize=(8,6))
```

```
# Wild type
plt.plot(common_time, avg_wt, "-", color="green", lw=2.5, label="WT Avg")
plt.fill_between(common_time, avg_wt - wt_std, avg_wt + wt_std, alpha=0.2, color="olive")

# Mutant
plt.plot(common_time2, avg_mut, "-", color="red", lw=2.5, label="EvD Avg")
plt.fill_between(common_time2, avg_mut - mut_std, avg_mut + mut_std, alpha=0.2, color="orange")

plt.xlabel("Time (h)", fontsize=24)
plt.ylabel(r"$CO_2$ Production rate (mL/min)", fontsize=24)
plt.xticks(fontsize=24)
plt.yticks(fontsize=24)
plt.title("AFM Wild Type vs EvD", fontsize=40)
plt.legend(fontsize=14, loc="upper left")
plt.tight_layout()
plt.show()
```

Script S3. Python script used for visualisation of production yields using violin plots. This script visualises the distribution of measured values (e.g., yield) across different samples. Data are extracted from an Excel file and reorganised into a structured format for plotting. Distributions are displayed using violin plots, with individual replicate measurements overlaid as jittered points. Mean values for each group are calculated and indicated on the plot as horizontal lines. This representation allows simultaneous assessment of central tendency, variability, and distribution shape across samples.

```
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

plt.rcParams['font.family'] = 'Tw Cen MT'
# === Read Excel File ===
file_path = r"file_path/file.xlsx"
df = pd.read_excel(file_path)
```

```
# Extract sample names and values
samples = df.iloc[:, 0]
values = df.iloc[:, 5]

# Create a new DataFrame
data = pd.DataFrame({
    "Sample": samples,
    "Value": values
})

# === Plot violin plot with dots ===
plt.figure(figsize=(8,6))

# Violin plot
sns.violinplot(x="Sample", y="Value", data=data, inner=None, cut=3, palette="Set2")

# Overlay dots (replicates)
sns.stripplot(x="Sample", y="Value", data=data, size=4, jitter=0.05, color="black")

grouped_data = data.groupby('Sample')['Value'].mean().reset_index()
for i, sample in enumerate(grouped_data['Sample']):
            mean_value = grouped_data[grouped_data['Sample'] == sample]['Value'].iloc[0]
     plt.hlines(y=mean_value, xmin=i - 0.25, xmax=i + 0.25, color='gray', linestyle='-', linewidth=2)
    print(sample, mean_value)

plt.xticks(rotation=0, ha="right", fontsize=24)
plt.yticks(ha="right", fontsize=24)
plt.xlabel("Strain", fontsize=24)
plt.ylabel("Yield", fontsize=24)
plt.title("EtOH Yield", fontsize=46)
```

```
plt.tight_layout()
plt.show()
```

Script S4. Python script used statistical comparison between two groups. The script calculates descriptive statistics (mean, SD, CV), tests normality (Shapiro–Wilk) and variance equality (Levene's test), and applies an appropriate significance test (Student's or Welch's t-test, or Mann–Whitney U test). Variability is further assessed via bootstrap confidence intervals of the coefficient of variation, and distribution differences are evaluated using the Kolmogorov–Smirnov test.

```
import pandas as pd
import numpy as np
from scipy import stats
import warnings
warnings.filterwarnings("ignore")

# === USER INPUT ===
file_path = r"file_path\file.xlsx"
df = pd.read_excel(file_path)
alpha = 0.05
n_boot = 10000  # for CV confidence interval

df.columns = ["Sample", "Val1", "Val2", "Val3", "Val4", "Yield"]
df = df.dropna(subset=["Sample", "Yield"])
df["Sample"] = df["Sample"].astype(str).str.strip()

groups = df["Sample"].unique()
if len(groups) != 2:
    raise ValueError(f"Expected 2 groups, found {groups}")
g1, g2 = groups
x1 = df.loc[df["Sample"] == g1, "Yield"].values
x2 = df.loc[df["Sample"] == g2, "Yield"].values

# === BASIC STATS ===
mean1, mean2 = np.mean(x1), np.mean(x2)
```

```
sd1, sd2 = np.std(x1, ddof=1), np.std(x2, ddof=1)
cv1, cv2 = sd1 / mean1, sd2 / mean2

print(f"\n=== BASIC STATS ===")
print(f"{g1}: n={len(x1)}, mean={mean1:.3f}, SD={sd1:.3f}, CV={cv1:.3f}")
print(f"{g2}: n={len(x2)}, mean={mean2:.3f}, SD={sd2:.3f}, CV={cv2:.3f}")

# === NORMALITY TESTS ===
sw1 = stats.shapiro(x1)
sw2 = stats.shapiro(x2)
normal = (sw1.pvalue > alpha) and (sw2.pvalue > alpha)
print("\n=== NORMALITY (Shapiro–Wilk) ===")
print(f"{g1}: p={sw1.pvalue:.4f}, {g2}: p={sw2.pvalue:.4f}")
print("- Data are normal" if normal else "→ Data are not normal")

# === VARIANCE EQUALITY ===
lev = stats.levene(x1, x2, center='median')
print(f"\n=== VARIANCE EQUALITY (Levene’s test) ===")
print(f"W={lev.statistic:.3f}, p={lev.pvalue:.4f}")

# === TEST FOR DIFFERENCE IN MEANS ===
if normal:
    if lev.pvalue > alpha:
        test = stats.ttest_ind(x1, x2, equal_var=True)
        test_used = "Student’s t-test (equal variances)"
    else:
        test = stats.ttest_ind(x1, x2, equal_var=False)
        test_used = "Welch’s t-test (unequal variances)"
else:
    test = stats.mannwhitneyu(x1, x2, alternative='two-sided')
    test_used = "Mann–Whitney U test (non-parametric)"

print(f"\n=== DIFFERENCE IN MEANS ===")
```

```
print(f"{test_used}: statistic={test.statistic:.3f}, p={test.pvalue:.4f}")
sig = "Significant" if test.pvalue < alpha else "Not significant"
print(f"Result: {sig} (p {('<' if test.pvalue<alpha else '≥')} {alpha})")

# === BOOTSTRAP CONFIDENCE INTERVAL FOR CVs ===
def bootstrap_cv(x, n_boot=10000):
    rng = np.random.default_rng()
    cvs = []
    for _ in range(n_boot):
        sample = rng.choice(x, len(x), replace=True)
        cvs.append(np.std(sample, ddof=1) / np.mean(sample))
    return np.percentile(cvs, [2.5, 97.5])

ci1 = bootstrap_cv(x1, n_boot)
ci2 = bootstrap_cv(x2, n_boot)

print(f"\n=== COEFFICIENT OF VARIATION (robustness) ===")
print(f"{g1}: CV={cv1:.3f} [95% CI {ci1[0]:.3f}–{ci1[1]:.3f}]")
print(f"{g2}: CV={cv2:.3f} [95% CI {ci2[0]:.3f}–{ci2[1]:.3f}]")

# === OPTIONAL: DISTRIBUTION DIFFERENCE (KS TEST) ===
ks = stats.ks_2samp(x1, x2)
print(f"\nKolmogorov–Smirnov test: D={ks.statistic:.3f},
p={ks.pvalue:.4f}")

# === SUMMARY ===
print("\n===== SUMMARY =====")
print(f"- Mean yield higher in {g2 if mean2>mean1 else g1} "
      f"({mean1:.3f} vs {mean2:.3f})")
print(f"- Variability lower in {g2 if cv2<cv1 else g1} "
      f"(CV {cv1:.3f} vs {cv2:.3f})")
if lev.pvalue > alpha:
    print("- No significant difference in variance (Levene p ≥ 0.05)")
```

```
else:
    print("- Variance difference significant (Levene p < 0.05)")
if ks.pvalue < alpha:
    print("- Overall distributions differ significantly (KS p < 0.05)")
else:
    print("- No major difference in overall distribution (KS p ≥ 0.05)")
```

Chapter 5

Insights and Future Perspectives

5.1 General overview: Refining *K. marxianus* for the circular bioeconomy

The transition from a petrochemical-based economy to a circular bioeconomy is not only an idealistic aspiration, but also a societal obligation. Throughout this thesis, some of the challenges inherent to this transition were explored, specifically focusing on the optimisation of microbial cell factories. While *S. cerevisiae* has long been the premier industrial yeast, in the last decade there has been a paradigm shift. The data and developments presented in the preceding chapters argue for a deeper and more diverse change towards non-conventional yeasts, with *K. marxianus* emerging as a protagonist of exceptional versatility.

This thesis aims at bridging the gap between the innate potential of *K. marxianus* as a host for bioprocesses (specifically because of its thermotolerance, rapid growth, and metabolic diversity) and its practical implementation in industrial biorefineries. To achieve this, a multi-layered approach was adopted: from setting up the theoretical frameworks of synthetic biology (Chapter 1) to fundamental physiological understanding of this yeast on a renewable feedstock (Chapter 2), through to the development of synthetic biology tools (Chapter 3), and finally, full industrial-scale application of a robust laboratory-evolved strain (Chapter 4). By unravelling *K. marxianus*' nitrogen sensing mechanisms, developing a standard secretion kit, and pushing its physiological limits through evolution, this thesis contributes to a holistic strategy for developing next-generation industrial bioprocesses.

5.2 Deciphering the nitrogen metabolic wiring of a non-WGD yeast

A central theme emerging from Chapter 2 was the realisation that bioprocesses cannot solely be optimised *in silico*. In industrial bioprocesses, particularly those utilising complex waste streams like whey permeate, nitrogen availability is frequently a rate-limiting factor specific of each substrate. A transcriptomic analysis of *K. marxianus* provided a snapshot of how this yeast perceives and responds to low availability of nitrogen in whey permeate, revealing evolutionary divergences from *S. cerevisiae*.

The comparison between *K. marxianus* and *S. cerevisiae* offers an interesting case study in evolutionary biology. As mentioned in Chapter 2, *S. cerevisiae* is the product of a hybridisation event, initially coined as a WGD (Gabaldón, 2020; Wolfe, 2015). The analysis relies on the hypothesis that this WGD event provided *S. cerevisiae* with regulatory redundancy, allowing for the sub-functionalisation of nutrient sensing pathways (Mattenberger et al., 2017). In contrast, *K. marxianus*, a pre-WGD yeast, appears to maintain a more rigid, ancestral wiring of its metabolic networks.

Overall, it was observed that *K. marxianus*, as *S. cerevisiae*, uses arginine pools to mobilise nitrogen towards anabolism, although some differences were also observed. Specifically, the data regarding methionine biosynthesis is of particular significance. These data suggest that, in *K. marxianus*, methionine acts as a rigid master switch for anabolic capacity. The strong upregulation of methionine biosynthetic genes in poor nitrogen conditions indicates a strict reliance on this amino acid as a proxy for cellular nitrogen status. This contrasts with the post-WGD landscape of *S. cerevisiae*, where the expansion of regulatory genes potentially allowed for the decoupling of specific anabolic signals from direct methionine sensing (Fang et al., 2023).

This observation is a rather logical extension of the WGD theory, that equipped *S. cerevisiae* with a greater set of sugar transporters or a second layer of regulatory proteins that allow this yeast to be an excellent alcohol-fermenting yeast (Donzella et al., 2023; Gabaldón, 2020; Purkanti & Thattai, 2022). The high redox price of methionine synthesis, combined with its connection to the SAM and folate cycles, makes this amino acid an ideal, although metabolically costly, sensor. In *K. marxianus*,

the inability to decouple growth signals from methionine levels may explain the higher expression of the methionine pathway under nitrogen starvation compared to *S. cerevisiae* (Rollero et al., 2019). This insight could have profound practical implications: it suggests that supplementing *K. marxianus* fermentations with sulphur-containing amino acids or engineering the redox balance of the sulphur assimilation pathway, could be a key strategy for unlocking higher growth rates in nutrient-poor media.

Other differences in other pathways were also observed: The expression of different amino acid oxidases (DAO genes), the purine degradation pathway or a cytoplasmic route to produce cysteine through *CYS2* and *cysB*; evidencing the distinct metabolism of *K. marxianus*.

The insights from Chapter 2 move beyond basic biology into process engineering. In the biorefinery context, one main goal is circularity. By using wastes and secondary products as feedstocks for microbial growth, the economic and ecologic cost of bioprocesses can be significantly reduced. Acknowledging that *K. marxianus* lacks the regulatory plasticity of *S. cerevisiae* means that process engineers must design new feeding strategies and solutions that maintain the ideal environment for this yeast to be used as microbial cell factory. The transcriptomic landscape defined here will serve as a map for rational media optimisation in a dairy biorefinery, moving away from an empirical approach towards a more rational data-driven formulation.

Nevertheless, it is important to highlight that high expression of biosynthetic genes does not strictly correlate with high pathway flux (Ben-Moshe et al., 2019; Petelski & Slavov, 2020), it can often indicate a compensatory mechanism for an inefficient enzyme or the presence of other layers of regulation. To confirm the hypothesis that methionine synthesis acts as a metabolic switch in *K. marxianus*, future work must move beyond RNA quantification to Metabolic Flux Analysis. By feeding the yeast with isotopically labelled nutrients the actual distribution of these atoms could be traced through the intracellular network. Specifically, ^{34}S and ^{15}N-labeling would be instrumental to resolve *K. marxianus* methionine metabolism. If this hypothesis is correct, a high flux of ^{15}N directing towards the methionine branch would be expected even under low nitrogen conditions, confirming a lack of flux redirection.

While isotope labelling remains central to define metabolic pathways, the use of genetically encoded biosensors could also shed light on this issue. By leveraging the transcriptome data from Chapter 2, new promoters responding to this environment could be developed and, together with the MoClo toolkit, *K. marxianus* strains expressing fluorescent sensors responsive to intracellular methionine or SAM levels could be engineered. This would enable the direct observation of the starvation response in live cells during culture, bridging the gap between the static data of the transcriptome and the dynamic reality of the fermenter.

5.3 A YTK-Compatible Platform for Secretion Engineering

While Chapter 2 highlighted the intrinsic differences between *K. marxianus* and *S. cerevisiae*, Chapter 3 provided solutions to bridge that gap. A recurring bottleneck in the adoption of non-conventional yeasts has been the lack of standardised genetic tools. Prior to this work, protein secretion in *K. marxianus* was hardly explored and often treated as a "black box," with only a few studies deepening in this yeast's particularities (Yarimizu et al., 2015). While the secretory capacity of this yeast is high, harnessing this trait was hindered by a lack of characterised secretion signals (Gombert et al., 2016; Tian et al., 2023; Zhou et al., 2018). By isolating secretion signals and characterising them using a YSD platform, this black box can be converted into a predictable engineering space.

The application of YSD in Chapter 3 served a dual purpose. First, it allowed for the rapid, high-throughput screening of secretion signals. Second, it increased the potential of *K. marxianus* as a platform for heterologous protein production. Moreover, the ability to display functional proteins on the cell surface at high temperatures could open new avenues for biotechnological applications, such as simultaneous saccharification and fermentation (SSF) or whole-cell biocatalysts that can operate under harsh industrial conditions.

Additionally, the characterisation of these secretion signals confirmed that secretion is not merely a function of the signal sequence but is context dependent: The interplay between the secretion signal, the target protein, and the host's folding machinery reinforces the need for a toolkit approach rather than an all-purpose

secretion signal. Thus, by offering a library of parts that can be screened for optimal performance, and releasing this toolkit to the community, this thesis helps reduce the limitations that hinder the adoption of *K. marxianus* as a production host. The decision to make this toolkit compatible with the YTK standard is strategic. It ensures that the *K. marxianus* secretion toolkit is not isolated in the synthetic biology landscape but is instead interoperable with libraries developed for other organisms.

While YSD offers high-throughput capability via flow cytometry, the results in Chapter 3 highlighted a critical limitation: the correlation between display efficiency and true secretion is not perfect. The tethering of the protein to the cell wall introduces steric constraints that do not exist for freely secreted proteins and the addition of Aga2p can also affect the final folding of target proteins. To fully unlock the secretory potential of *K. marxianus*, future work must look towards the application of screening technologies that physically isolate single cells and/or their secretion products, bypassing the biophysical limitations of cell-surface tethering (Cleaver et al., 2025; Kishkevich et al., 2025; Lebrun et al., 2023; Wei et al., 2023).

Looking forward, the secretion toolkit developed here is a finite resource comprising only a few native *K. marxianus* secretion signals. However, it has the potential to be dynamic thanks to the modularity of the MoClo system. With the advent of machine learning models like AlphaFold and protein language models, it would be possible to predict secretion signal efficiency *in silico*. Future work should focus on generating training datasets using the YSD platform to feed these models with signals that fill the full DoE space. Eventually, it would be possible to move from screening a library of 10 secretion signals to computationally generating the ideal one for a specific protein-host combination.

Moreover, an expansion of the *K. marxianus* MoClo toolkit to include parts from the pangenome is essential. As shown in Chapter 2, there is significant strain-level variation in *K. marxianus*; mining diverse isolates for novel promoters, selection markers, origins of replication, and terminators will enrich the synthetic biology toolbox, allowing for fine-tuned control of gene expression and increased metabolic capacity.

Finally, the rapid evolution of *de novo* DNA synthesis technologies must be considered. Currently, modular standards like YTK are indispensable because they circumvent the high cost of synthesising long DNA fragments; they allow researchers to piece together parts like Lego blocks. However, the emerging capability to generate full DNA fragments cost-effectively will render physical assembly standards obsolete. In the near future, the constraint of part compatibility (e.g. removing internal restriction sites) will disappear. Instead of assembling a plasmid from physical parts, researchers will move toward a workflow of pure digital design followed by the direct synthesis of entire expression cassettes. This shift will also help resolving the structural compromises currently required by standardisation. As demonstrated in Chapter 3, the design of the Yeast Toolkit imposes a Gly–Ser link between the secretion signal and the target protein, which was found to affect secretion efficiency. While a seamless junction for the Mel1p case was successfully engineered, it came at the cost of breaking YTK compatibility. In this assembly-free future, toolkits will transition from a physical freezer box of plasmids to a digital repository of characterised elements, accelerating the DBTL cycle.

Ultimately, the establishment of this standardised secretion platform enables a strategic pivot in the biorefinery value chain: moving from bulk commodities to high-value bioproducts. While Chapter 4 demonstrated the efficient conversion of whey permeate into bioethanol, the economic margins for this sector remain tight. By coupling the secretion toolkit with the organism's innate capacity to metabolise lactose, the potential to transform agro-industrial side-streams directly into high added value recombinant proteins is unlocked. In this model, whey permeate becomes a functional media for the secretion of industrial enzymes, food proteins or therapeuticals. Furthermore, the QPS and GRAS status of *K. marxianus* significantly lowers the regulatory barrier for deploying these products, offering a distinct advantage over other microbial hosts.

5.4 Achieving industrial robustness through ALE

While Chapter 3 demonstrated the precision of rational design, Chapter 4 employed a broader approach based on ALE. The development of an ethanol tolerant strain

capable of robust valorisation of whey permeate demonstrates the extraordinary power of selection.

The focus on whey permeate in Chapter 4 directly addresses the sustainability goals outlined in Chapter 1. Whey permeate is a high-volume, high-BOD (Biological Oxygen Demand) pollutant if released into the environment. Its conversion into bioethanol transforms a liability into an asset. The success of bringing this process to industrial levels validates *K. marxianus* as a superior host for lactose-based biorefineries. Publicly available performance data from trials of this magnitude (120,000 L) remains exceptionally rare. Encouraging such transparency in industry-academia collaborations is essential to accelerate the translation of biotechnological discoveries into tangible bioprocesses.

It is well-established that *S. cerevisiae* is naturally a more efficient ethanol producer, owing to its specialised regulatory network and strong Crabtree effect (Hagman et al., 2014). *K. marxianus*, lacking this specific regulatory architecture, does not intrinsically prioritise fermentation to the same degree (Malina et al., 2021). However, the selection of *K. marxianus* for this application attends to two decisive factors: its native capacity to metabolise lactose, and its superior industrial robustness traits, including thermotolerance and rapid growth (Bilal et al., 2022; De Micco et al., 2025). Traits like ethanol tolerance or osmotic stress resistance are polygenic, involving complex remodelling of the membrane, chaperones, and central metabolism (Dhar et al., 2011; Lertwattanasakul et al., 2023). ALE allowed the organism to find its own solution to these multifactorial problems. The resulting strain, which outperformed the wild type in the industrial environment of Carbery, reveals a crucial distinction in the nature of *K. marxianus*: while Chapter 2 highlighted the rigidity of its immediate regulatory wiring, Chapter 4 demonstrated the extraordinary adaptive potential of its genome over time.

Moreover, in the context of the European food landscape, the regulatory status of the microbial host is often a decisive factor. While the tools developed in Chapter 3 allow for precise manipulation, the deployment of GMOs in the food and feed sectors remains heavily regulated and consumer sensitive. This reality was a primary constraint in the collaboration with Carbery. By utilising ALE, a non-GMO strain was

obtained showing robust performance in the industrial set-up that can be applied in Carbery's production process without compliance issues.

Finally, the work in Chapter 4 on whey permeate valorisation represents a concrete step forward. However, while ALE currently serves as an effective tool for navigating complex landscapes and outputs, it remains a black box approach. As we integrate AI and systems biology to fully decipher cellular regulation, we will eventually move beyond empirical tools. The ultimate goal is to decipher the specific genetic variants that evolution selects for and catalogue them, essentially converting evolutionary outcomes into engineering inputs. This will unlock an *à la carte* design paradigm: treating *K. marxianus* as a universal chassis where specific genetic modules can be swapped in like parts of a machine. This convergence of evolutionary discovery and rational assembly will move the bioeconomy towards a redistributed manufacturing model, where a single robust chassis can be rapidly reconfigured to valorise any local waste stream.

5.5 Synthesis

The contrasts between the chapters in this thesis highlight a dichotomy in modern biotechnology: **Rational Design** (GG Cloning) versus **Evolutionary Engineering** (ALE).

- **Evolutionary Engineering (Chapter 4)** offers robustness and global regulatory network remodelling that is beyond current predictive capabilities.
- **Rational Design (Chapter 3)** offers speed, predictability, and the ability to introduce heterologous traits (like novel pathways) that evolution would never reach on its own.
- **Omics (Chapter 2)** serves as the bridge, explaining the reasons behind the observed phenotypes.

However, as introduced in Chapter 1, the future of microbial engineering lies in the integration of these approaches. For instance, the transcriptomic insights from Chapter 2 can now be used to reverse-engineer the ALE strains from Chapter 4. By sequencing the evolved strains and comparing their expression profiles to the baseline data generated in Chapter 2, the specific mutations that confer high ethanol

production in poor nitrogen feedstocks can be identified. Once identified, these mutations can be reintroduced into the clean, standardised backgrounds developed in Chapter 3 using CRISPR/Cas9 or applying directed evolution.

This recreates the iterative cycle described in Chapter 1: **Learn** (Transcriptomics), **Design** (SynBio/ALE), **Build** (Strain development), **Test** (Fermentation) and **Learn** again.

5.7 Concluding Remarks

This PhD thesis has journeyed through the landscape of *K. marxianus*, delivering tools and data to help transform this yeast from a promising microbe into a refined chassis for the bioeconomy. Its metabolism was mapped, the engineering tools built, and it was evolved to withstand the rigors of industry.

The discussion in Chapter 1 is not about imposing human will upon biology, but rather about understanding its logic, such as the nitrogen sensing constraints found in Chapter 2, and working within those constraints to achieve sustainable outcomes. Whether through the precise excision of DNA in Chapter 3 or the gradual pressure of selection in Chapter 4, the goal remains the same: to create an interface between biological capability and industrial necessity: the yeast.

As we stand on the precipice of a climate-defined century, the tools and strains developed in this work offer a tangible contribution to the arsenal of the circular economy. *K. marxianus* is no longer just the "fastest-growing eukaryote"; it is now a programmable, robust, and understood platform, ready to take its place alongside *S. cerevisiae* as a pillar of sustainable manufacturing.

References

Ben-Moshe, S., Shapira, Y., Moor, A. E., Manco, R., Veg, T., Bahar Halpern, K., & Itzkovitz, S. (2019). Spatial sorting enables comprehensive characterization of liver zonation. *Nature Metabolism 2019 1:9*, *1*(9), 899–911. https://doi.org/10.1038/s42255-019-0109-9

Bilal, M., Ji, L., Xu, Y., Xu, S., Lin, Y., Iqbal, H. M. N., & Cheng, H. (2022). Bioprospecting Kluyveromyces marxianus as a Robust Host for Industrial Biotechnology. *Frontiers in Bioengineering and Biotechnology*, *10*, 851768. https://doi.org/10.3389/FBIOE.2022.851768

Cleaver, A., Luo, R., Smith, O. B., Murphy, L., Schwessinger, B., & Brock, J. (2025). High-throughput optimisation of protein secretion in yeast via an engineered biosensor. *Trends in Biotechnology*, *43*(4), 838–867. https://doi.org/10.1016/J.TIBTECH.2024.11.010

De Micco, C., Radice, M., Di Fidio, N., Agrimi, G., & Pisano, I. (2025). Physiological characterization of Kluyveromyces marxianus strains for the development of a microbial platform to obtain bioethanol from cheese whey permeate. *Frontiers in Microbiology*, *16*, 1663736. https://doi.org/10.3389/FMICB.2025.1663736

Dhar, R., Sägesser, R., Weikert, C., Yuan, J., & Wagner, A. (2011). Adaptation of Saccharomyces cerevisiae to saline stress through laboratory evolution. *Journal of Evolutionary Biology*, *24*(5), 1135–1153. https://doi.org/10.1111/J.1420-9101.2011.02249.X

Donzella, L., Sousa, M. J., & Morrissey, J. P. (2023). Evolution and functional diversification of yeast sugar transporters. *Essays in Biochemistry*, *67*(5), 811–827. https://doi.org/10.1042/EBC20220233

Fang, W., Jiang, L., Zhu, Y., Yang, S., Qiu, H., Cheng, J., Liang, Q., Tu, Z. cai, & Ye, C. (2023). Methionine restriction constrains lipoylation and activates mitochondria for nitrogenic synthesis of amino acids. *Nature Communications 2023 14:1*, *14*(1), 2504-. https://doi.org/10.1038/s41467-023-38289-9

Gabaldón, T. (2020). Hybridization and the origin of new yeast lineages. *FEMS Yeast Research*, *20*(5), 40. https://doi.org/10.1093/FEMSYR/FOAA040

Gombert, A. K., Madeira, J. V., Cerdán, M. E., & González-Siso, M. I. (2016). Kluyveromyces marxianus as a host for heterologous protein synthesis. *Applied Microbiology and Biotechnology*, *100*(14), 6193–6208. https://doi.org/10.1007/s00253-016-7645-y

Hagman, A., Säll, T., & Piškur, J. (2014). Analysis of the yeast short-term Crabtree effect and its origin. *FEBS Journal*, *281*(21), 4805–4814. https://doi.org/10.1111/febs.13019

Kishkevich, A., Ciurkot, K., & Ellis, T. (2025). YTK Display-and-Secrete: Screening for Optimal Protein Secretion Elements in Saccharomyces cerevisiae. *ACS Synthetic Biology*. https://doi.org/10.1021/ACSSYNBIO.5C00264

Lebrun, E., Shenshin, V., Plaire, C., Vigneres, V., Pizette, T., Dumas, B., Nicaud, J.-M., & Mottet, G. (2023). Efficient full-length IgG secretion and sorting from single yeast clones in droplet picoreactors. *Lab on a Chip*, *23*, 3487–3500. https://doi.org/10.1039/D3LC00403A

Lertwattanasakul, N., Pattanakittivorakul, S., Nitiyon, S., Matsutani, M., Oguchi, A., Hirata, K., Kosaka, T., Limtong, S., & Yamada, M. (2023). Mutants with Enhanced Multi-Stress Tolerance of Kluyveromyces marxianus and Their Ability for Ethanol Fermentation. *Fuels 2023, Vol. 4, Pages 469-483*, *4*(4), 469–483. https://doi.org/10.3390/FUELS4040029

Malina, C., Yu, R., Bjorkeroth, J., Kerkhoven, E. J., & Nielsen, J. (2021). Adaptations in metabolism and protein translation give rise to the Crabtree effect in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(51), e2112836118. https://doi.org/10.1073/PNAS.2112836118

Mattenberger, F., Sabater-Muñoz, B., Toft, C., & Fares, M. A. (2017). The phenotypic plasticity of duplicated genes in Saccharomyces cerevisiae and the origin of adaptations. *G3: Genes, Genomes, Genetics*, *7*(1), 63–75. https://doi.org/10.1534/G3.116.035329

Petelski, A. A., & Slavov, N. (2020). Analyzing Ribosome Remodeling in Health and Disease. *Proteomics*, *20*(17–18), 2000039. https://doi.org/10.1002/PMIC.202000039

Purkanti, R., & Thattai, M. (2022). Genome doubling enabled the expansion of yeast vesicle traffic pathways. *Scientific Reports*, *12*(1), 1–11. https://doi.org/10.1038/S41598-022-15419-9

Rollero, S., Bloem, A., Ortiz-Julien, A., Bauer, F. F., Camarasa, C., & Divol, B. (2019). A comparison of the nitrogen metabolic networks of Kluyveromyces marxianus and Saccharomyces cerevisiae. *Environmental Microbiology*, *21*(11), 4076–4091. https://doi.org/10.1111/1462-2920.14756

Tian, T., Wu, X., Wu, P., Lu, X., Wang, Q., Lin, Y., Liu, C., Zhou, J., Yu, Y., & Lu, H. (2023). High-level expression of leghemoglobin in Kluyveromyces marxianus by remodeling the heme metabolism pathway. *Frontiers in Bioengineering and Biotechnology*, *11*, 1329016. https://doi.org/10.3389/FBIOE.2023.1329016

Wei, X., Penkauskas, T., Reiner, J. E., Kennard, C., Uline, M. J., Wang, Q., Li, S., Aksimentiev, A., Robertson, J. W. F., & Liu, C. (2023). Engineering Biological Nanopore Approaches toward Protein Sequencing. *ACS Nano*, *17*(17), 16369–16395. https://doi.org/10.1021/ACSNANO.3C05628

Wolfe, K. H. (2015). Origin of the Yeast Whole-Genome Duplication. *PLOS Biology*, *13*(8), e1002221. https://doi.org/10.1371/JOURNAL.PBIO.1002221

Yarimizu, T., Nakamura, M., Hoshida, H., & Akada, R. (2015). Screening of accurate clones for gene synthesis in yeast. *Journal of Bioscience and Bioengineering*, *119*(3), 251–259. https://doi.org/10.1016/J.JBIOSC.2014.08.006

Zhou, J., Zhu, P., Hu, X., Lu, H., & Yu, Y. (2018). Improved secretory expression of lignocellulolytic enzymes in Kluyveromyces marxianus by promoter and signal sequence engineering. *Biotechnology for Biofuels 2018 11:1*, *11*(1), 235-. https://doi.org/10.1186/S13068-018-1232-7

Acknowledgements

I want to start by expressing my gratitude to Taighde Éireann that funded this PhD thesis within the 2021 Government of Ireland Postgraduate Scholarship Grant with project number GOIPG/2021/435.

I sincerely thank my supervisors Prof. John Morrissey and Prof. Maria de Sousa Gallagher for allowing me to pursue a PhD and carry out the research in this thesis, and the opportunities this PhD has provided me. Thank you, John, for the opportunity you gave me in Ireland and for keeping me even if I blew up the heaters the first week.

Primero quiero agradecer eternamente a mi esposa, amore y profunda compañera Lorena. Sin ti jamás podría haber llegado a donde estoy. Este doctorado no ha sido solo una tesis, sino un viaje que nos ha unido y transformado. Gracias por viajar junto a mí y volverme mejor persona, aguantar lo pesado que soy y no explotar después de pedirte ayuda y luego no esuchar. Te amo por siempre.

Quiero agradecer también a mis padres, Vicen y Rafa, por haber entendido mis motivaciones para abandonarles en Valencia y no solo no oponerse sino alentarme. Sé que ha sido y seguirá siendo duro. Gracias, os quiero mucho papás.

También quiero agradecer a mi hermano Xavi, Ana y al pequeñín Einar (y a Viggo). Yo habré completado un doctorado, pero vosotros habéis completado algo mucho más valioso.

Finalmente al resto de mi familia, mis tíos, mis iaios i els meus amics a casa, por siempre quererme y esperarme con la misma sonrisa a que vuelva después de meses sin verme el pelo, os quiero a todos.

Un altro grande baccio a tutta la mia nuova famiglia italiana, che anche essendo nuovo tra di loro dalla prima giornata mi hanno trattato come se fosse nato a Taranto. Grazie e vi Amo.

I want to thank also all the new friends I made along these six years I have been living in Cork. You all helped me pushing this thesis forward through the rainy days in one way or the other: by having a coffee or a chat with me, joining an Italian party or watching Sarremo. Especially, Roberto and Monica, thank you for the intense debate

about all my stupid questions such us what the best ice cream flavour in the universe is in absolute or which are the five most populated countries in South America. Thank you all.

I thank Prof. Tom Ellis for the opportunity to stay in his laboratory and to Stacey for the help and guidance during that period in London. I also want to extend my gratitude to Yeast4Bio COST action (CA18229) and to the Microbiology Society for partially funding my stay in London.

Many thanks to Carbery Food Ingredients staff in particular to Pat Mulcahy and Aine Hallihan who I closely collaborated with during the past years and let me learn a lot about how real bioprocesses operate.

I want to extend my gratitude also to all the staff in the School of Microbiology and SUSFERM and all the other people that helped me in one way or another to complete this thesis. They all are brilliant, but I cannot finish this acknowledgement without highlighting the role of the incredible Dr. John O'Callaghan, thanks for all the help, I hope I never got too annoying.