On metabolic and phenotypic diversity in yeast

Stephan Kamrad

The Molecular Biology of Metabolism Laboratory, The Francis Crick Institute

and

Department of Genetics, Evolution and Environment, University College London

A thesis submitted for the degree of

Doctor of Philosophy

University College London.

December 2020

I, Stephan Kamrad, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

This thesis explores metabolic and phenotypic diversity in the two model yeasts *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae*. Colony screens are a classical and powerful technique for investigating these topics, but there is a lack of modern, scalable bioinformatics tools. To address this need, I have developed *pyphe* which greatly facilitates colony screen data acquisition and statistical analysis. I explore optimal experimental designs, especially regarding the usefulness of timecourse imaging and colony viability analysis. *Pyphe* is used in a functional genomics screen, aiming to find functions for a set of largely uncharacterised lincRNAs. We identify hundreds of new lincRNA-associated phenotypes across numerous conditions and compare lincRNA phenotype profiles to those of coding-gene mutants. Next, I have used *pyphe* to investigate the respiration/fermentation balance of wild *S. pombe* isolates. Contrary to the expectation that glucose completely represses respiration in this Crabtree-positive species, I find that strains generally strike a balance and that individual strains differ significantly in their residual respiration activity. This is associated with an unusual miss-sense variant in *S. pombe's* sole pyruvate kinase gene. Its impact is dissected in detail, revealing a change in flux through pyruvate kinase and associated changes in gene expression, metabolism, growth and stress resistance. Finally, I explore how extracellular amino acids interact with cellular metabolism, with the aim of answering the important question whether or not clonal yeast cultures segregate into heterogeneous producer/consumer populations that exchange amino acids. I develop a novel proteomics-based method that characterises amino acid labelling patterns in peptides. I find that the supplementation of some, but not all amino acids completely suppresses self-synthesis. However, I find no evidence for heterogeneous responses of our laboratory *S. cerevisiae* strain, but the functionality of the method is demonstrated clearly. Overall, this work represents several advancements to our understanding of yeast metabolism and physiology, as well as new experimental and computational methods.

Impact Statement

This thesis explores fundamental biological questions and its findings will impact basic and applied research in these areas. Primarily, it contains new data, conclusions and concepts relating to yeast physiology, metabolism and systems biology, which will have an impact through their publication in academic journals. Specifically, I have characterised how yeast cells respond to the availability of extracellular amino acids which shines new light on the role and behaviour of cells in nutrient-rich environments (Chapter 3, manuscript under preparation). Additionally, we have uncovered numerous novel lincRNA-associated phenotypes and produced a large-scale systematic dataset which will provide a valuable resource for future targeted investigations of lincRNA function (Chapter 1, manuscript under preparation). Furthermore, I have shown that a mutation in the glycolytic enzyme *pyk1*, which is at first sight deleterious and lowers flux, can have a protective effect against oxidative stress in *S. pombe* wild strains (Kamrad et al., 2020a). This adds to a body of work linking metabolism to stress resistance and might stimulate further work considering this as an important evolutionary mechanism. I have also shown that specific amino acids accumulate to very high intracellular levels when supplemented in the media (Olin-Sandoval et al., 2019) and in the case of lysine, this 'harvesting' is a novel stress resistance mechanism against oxidants. This raises questions about the physiological impact of harvesting other amino acids, which could be a promising starting point of future studies. Measurements performed by me have also revealed amino acid imbalances in *S. pombe* cells upon inhibition of respiration, proving a novel a link between these two key metabolic pathways in this species (Malecki et al., 2020). This publication and Kamrad et al. (2020a) together are two puzzle pieces exploring why *S. pombe* appears to be more reliant on respiration than *S. cerevisiae*, an important difference in the physiology of these two key model species. Finally, I have contributed to a review article discussing strategies for multi-omic experiments and data analysis (Haas et al., 2017).

Independent of the biological conclusions, my work provides two key methodological advancements. I have published *pyphe*, a fully documented, open-source collection of command line tools for building data analysis pipelines for colony fitness screens. Besides being widely used in the Bähler and Ralser groups, I have already heard from outside users, and I hope that *pyphe* will be adopted widely. I have shown that colony viability staining with phloxine B can provide an orthogonal and independent readout in colony fitness screens (Kamrad et al., 2020b). This raises important follow-up questions of how social microbes like yeast grow in colonies and how the colony growth phenotype is determined by the ensemble of individual level. I have also developed a conceptually new method for detecting the heterogeneous utilisation of amino acids (Chapter 3). This method will be used by us and, once published, by others to deepen our understanding of metabolic single-cell heterogeneity.

The concepts, results and methods describing metabolic single cell heterogeneity could have important clinical implications if they are transferable to pathogenic microbes such as the yeast *Candida albicans*. It has been speculated that metabolic heterogeneity could influence antibiotic resistance and persistence, and this could be further explored using my methods. Eventually, drug or drug combinations could suppress metabolic heterogeneity thereby increasing the effectiveness of antibiotics.

References

Kamrad*, S., Grossbach*, J., Rodríguez-López, M., Mülleder, M., Townsend, S., Cappelletti, V., Stojanovski, G., Correia-Melo, C., Picotti, P., Beyer, A., Ralser, M., Bähler, J. (2020a). Pyruvate kinase variant of fission yeast tunes carbon metabolism, cell regulation, growth and stress resistance. Molecular Systems Biology 16: e9270

Kamrad, S., Rodríguez-López, M., Cotobal, C., Correia-Melo, C., Ralser, M. and Bähler, J. (2020b). Pyphe, a python toolbox for assessing microbial growth and cell viability in high-throughput colony screens. eLife 9: e55160

Malecki, M., **Kamrad, S.**, Ralser, M. and Bähler, J. (2020). Mitochondrial respiration is required to provide amino acids during fermentative proliferation of fission yeast. EMBO reports 21: e50845

Olin-Sandoval, V., Yu, J. S. L., Miller-Fleming, L., Alam, M. T., **Kamrad, S.**, Correia-Melo, C., Haas, R., Segal, J., Navarro, D. A. P., Herrera-Dominguez, L. et al. (2019). Lysine harvesting is an antioxidant strategy and triggers underground polyamine metabolism. Nature 572: 249–253

Haas, R., Zelezniak, A., Iacovacci, J., **Kamrad, S.**, Townsend, S. and Ralser, M. (2017). Designing and interpreting 'multi-omic' experiments that may change our understanding of biology. Current Opinion in Systems Biology 6: 37–45

Acknowledgements

I am immensely grateful for the continuous support, trust and mentorship I received from both my supervisors Markus Ralser and Jürg Bähler. It was a privilege to be a member of their two labs which are exceptionally friendly and collaborative environments. Specifically, I am grateful for the many discussions with the Ralser lab team working on microbial communities, Clara Correia-Melo, Jason Yu, Benjamin Heinecke, Lucía Herrera-Dominguez, Susann Zilkenat and Johannes Hartl. For specific projects, I worked closely with María Rodríguez-López, Simran Aulakh, Catalina Romana, Shajahan Anver, Clara Correia-Melo, Michal Malecki, Jan Grossbach and Michael Mülleder, and learnt a lot in the process. Christoph Messner, Michael Mülleder, Matt White and Spyros Vernadis have taught me everything I know about mass spectrometry. I learnt a lot about automatisation and high-throughput techniques from John Townsend whom I also thank for many stimulating discussions. I would like to thank the lab managers of both groups, Mimoza Hoti, Lukasz Szyrwiel, Enrica Calvani and Cristina Cotobal for their invaluable work and countless favours.

My thesis committee, Nataša Pržulj, Richard Mott and Nicholas Luscombe always provided valuable feedback and guidance. Many thanks also to the Crick Science Support, especially Robert Moore, Jade Smith-Hall and Muhammed Miah. Furthermore, I would like to acknowledge the following scientists: Stephan Noack for sharing data from his 2012 publication, Doug Speed for help with GWAS and heritability analysis, Jamie Macpherson for advice on PYK activity assays, Daniel Jeffares for advising on *S. pombe* wild strains, David Ellis for valuable discussion and critical reading of manuscripts, Quentin Saintain for useful feedback on beta versions of *pyphe*.

I am grateful to my family and to Rowan Henry, Caroline Gee, Joe Halstead, Henry Fried, Louisa Birkner and Leonie Köster for their love and support.

Contents

List of Figures

Preface

As a member of two groups with complementary expertise, I was exposed to a wide range of current questions and methods. While both groups use high-throughput and systems-scale approaches, the Bähler laboratory focuses mainly on genetics and transcriptional regulation while the Ralser laboratory investigates the regulation of metabolism and of the proteome. The work described in this thesis therefore spans all these omic-domains, aiming for an integrated understanding of cells as multi-layered, complex biological systems. Chapter 1 investigates the relationship between individual genes (in this case lincRNA genes) and the phenotype. Knock-out mutants are systematically assayed in a functional genomics workflow, in the hope of uncovering novel connections that might indicate gene functions. Inversely, Chapter 2 begins with an interesting phenotype (the respiration/fermentation balance), the genetic basis of which is interrogated using population genomics. Having identified a linked mutation, a multi-omics investigation of the transcriptome, proteome, metabolome and phenome is employed to characterise its physiological impact. Finally, Chapter 3 focuses on the interplay between the availability of extracellular nutrients and cellular metabolism, which is crucial in determining how cells interact with their surrounding niche.

In order to interrogate these different cellular processes and entities, all three projects described in this thesis employ interdisciplinary approaches and combine dry and wet work. I was trained initially in biochemistry and then later also in modelling and machine learning. In the Ralser group, I also learnt to use different mass spectrometry techniques for proteomics and metabolomics. Chapter 1 describes a novel, accessible and fully documented computational data analysis pipeline applied to several relevant biological datasets. Chapter 2 starts with a genome-wide association study performed on a collection of *S. pombe* wild strains, the results of which are further explored with a range of classical microbiological and biochemical approaches such as enzyme assays, oxygen consumption measurements

and growth curves. Finally, Chapter 3 uses a combination of proteomics and modelling, which only together allow for a quantitative understanding of the observed phenomena.

I massively enjoyed being part of collaborative projects and was fortunate to work with many fantastic colleagues. The work forming the basis of Chapter 1 was done in close collaboration with María Rodríguez-López, Cristina Cotobal and Shajahan Anver who collected key phenotypic data which I then analysed. Chapter 2 describes a project done in close collaboration with the Andreas Beyer group from the University of Cologne. Jan Grossbach analysed the two gene expression datasets which form a key part of the study. For these collaborative projects (which do not include Chapter 3), this thesis focuses on my contributions. Unless otherwise stated, data were collected, analysed and visualised by me and the work of others was only included if required for the flow of scientific arguments. The contributions of other individual researchers are clearly outlined in this thesis in the appropriate places (usually figure legends).

Chapter 1

High-throughput colony screening with pyphe identifies novel lincRNA-associated phenotypes

Eukaryotic genomes are pervasively transcribed and contain numerous long non-coding RNAs (lncRNAs). To date, the vast majority of intergenic lncRNAs (lincRNAs) in the model yeast *Schizosaccharomyces pombe* have no known biological roles. This project uses a functional genomics approach to generate novel hypotheses about lincRNA function. A library of lincRNA knock-out mutants is subjected to a wide range of growth conditions and fitness is assessed based on characteristics of colonies. LincRNA knock-out mutants show no growth effect in the majority of conditions, and when so only a subtle one, requiring an experimental setup and data analysis approach that simultaneously achieves high throughput and high precision. To this end, a new data analysis pipeline called *pyphe* is developed. *Pyphe* is made available freely, as a fully documented collection of command line tools written in python. It builds on current state-of-the-art normalisation techniques and implements the use of colony redness scores obtained by staining with phloxine B. Colony redness is shown to be closely related to the fraction of viable cells in the colony and to provide an orthogonal and independent readout to classical colony size screens. Applied to the lincRNA knock-out library, *pyphe* revealed 351 novel lincRNA associated phenotypes, covering half the deletion collection. Integrating phenotype profiles of lincRNA genes with that of coding genes reveals co-clustering of members of both gene sets, suggesting novel functional links.

The method development part of this project has been published in *eLife* (Kamrad et al., 2020b) and a manuscript describing the biological results is currently being finalised.

This chapter follows the arguments laid out in these manuscripts, while often providing substantial additional information and explanations. The lincRNA knock-out library was created by María Rodríguez-López, Cristinal Cotobal and others. The laboratory work for collecting the lincRNA phenome dataset was done by María Rodríguez-López, Cristinal Cotobal and Mimoza Hoti, the other datasets illustrating the method were collected by me. Clara Correia-Melo and I performed the described flow cytometry measurements together. All plots shown in this chapter were results of my own analyses, except in the rare cases where this is explicitly stated.

1.1 Introduction

1.1.1 The biological function of lincRNAs is poorly understood

Around the turn of the millennium, the publication of the genome sequences of many key eukaryotic model species, including the yeasts *Saccharomyces cerevisiae* (Goffeau et al., 1996) and *Schizosaccharomyces pombe* (Wood et al., 2002), marked the beginning of a new era in biological research (Giani et al., 2020). An important, difficult and ongoing challenge has been the systematic characterisation and annotation of functional units contained in them. Despite great efforts, a relatively large number of genes, approximately 20% in both *S. cerevisiae* and *S. pombe* (Wood et al., 2019), still have no known biological role to date. Simultaneously, coding regions make up only a fraction of eukaryotic genomes. Non-coding elements can be functionally diverse and include telomeric and centromeric regions serving a structural role, regulatory binding sites for transcriptional control, untranslated regions of protein-coding genes (UTRs and introns), as well as regions encoding functional RNA molecules (which are transcribed but not translated). The last include key components of the translational machinery: transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs). More recently, deep RNA sequencing has enabled the detecting countless new ncRNAs virtually all over the genome (Atkinson et al., 2012), making up a substantial fraction of the transcriptome in *S. pombe* (Marguerat et al., 2012; Atkinson et al., 2018) and *S. cerevisiae* (David et al., 2006; Xu et al., 2009). The repertoire of ncRNAs in animals, and their potential role in cancer, is a vast field outside the scope of this thesis (Slack and Chinnaiyan, 2019). NcRNAs are categorised based on their function, length, location in the genome and their breakdown pathway. NcRNAs more than 200 nucleotides in length are as classified as long non-coding RNAs (lncRNAs). Different groups of lncRNAs are degraded by different pathways: Cryptic unstable transcripts (CUTs) are degraded by the

nuclear RNA exosome, Xrn1-sensitive unstable transcripts (XUTs) are degraded by Xrn1 in the cytoplasm, and Dicer-sensitive unstable transcripts (DUTs) are dependent on the RNAi machinery (Atkinson et al., 2018). LncRNAs are found in different positions relative to coding genes. For example anti-sense ncRNAs are located in the same position but are oriented oppositely to coding genes and have been described to regulate their transcription (Wery et al., 2018), while bi-directional promoters produce ncRNAs starting at the same promoter as a coding transcript but proceeding in the opposite direction (Xu et al., 2009). However, many lncRNAs have their own transcription start sites and can be classed as intergenic or overlapping depending on their relative position to protein-coding genes. LincRNAs are the focus of this project as they represent physically distinct, independent units which can be mutagenised without directly affecting protein-coding genes.

The pervasiveness and low expression of many lncRNAs under standard laboratory conditions had initially raised doubts over their biological relevance. For example, Marguerat et al. (2012) have found the majority of lncRNAs to be expressed at less than a single copy per cell in *S. pombe*. With increasingly deep sequencing, there is a danger of capturing 'transcriptional noise', ie essentially errors of the transcriptional machinery. However, several lines of evidence now indicate that at least some lncRNAs indeed have specific cellular roles. Those include primarily their evolutionary conservation (Ponjavic et al., 2007) and an increasing number of focused studies dissecting the role of individual lncRNAs. The most highly transcribed lncRNAs show evidence for purifying selection in *S. pombe* wild strains (Jeffares et al., 2015) and transposon insertions in many lncRNAs result in detectable fitness defects (Grech et al., 2019). The function and mechanism of several individual fungal lncRNAs have been described in detail (reviewed in Till et al. (2018)) and were found to regulate diverse cellular processes such as meiosis, metabolism and stress resistance via different mechanisms. However, with at least 7000 lncRNAs described in *S. pombe* to date (Atkinson et al., 2018), a systematic, unbiased overview of lncRNA function is required to determine which lncRNAs are functional and in which context. One approach has been to deeply sequence the transcriptome in specific conditions or physiological states, looking for patterns of expression changes of lncRNAs. Besides revealing novel lncRNAs, this approach can generate hypotheses about the functional roles of lncRNAs, eg based on the observation that many appear to be strongly upregulated in meiosis (Atkinson et al., 2018). However, this indirect approach is unable to causally link lncRNAs to phenotypes

and functional roles. Here, for protein-coding genes, the availability of knock-out libraries has been the basis for systematic reverse genetics approaches to deciphering gene function, known as functional genomics (Giaever and Nislow, 2014; Brochado and Typas, 2013). In *S. cerevisiae*, a knock-out library comprising tRNAs, small ncRNAs and lncRNAs has been created (Parker et al., 2017) using the traditional method of homologous recombination with a resistance cassette. This library has been screened for phenotypes (Parker et al., 2018; Balarezo-Cisneros et al., 2020) and the transcriptome of a small set of knock-outs has been analysed (Balarezo-Cisneros et al., 2020), providing valuable clues and starting points for focused investigations. In particular, these studies have identified new essential ncRNAs and numerous condition specific fitness effects in heterozygous and haploid mutants. However, studies investigating lncRNA function using this library are limited by the number of lncRNA mutants (more than half of the 443 knocked-out genes are tRNAs, another 65 are small ncRNAs) and the method by which mutants were constructed. The kanMX4 cassette which is left in the genome in the knocked-out locus contains the strong TEF promoter which can significantly alter the chromatin state in the region, affecting the transcription of nearby genes, potentially leading to artefacts.

Addressing these issues, the Bähler group has created a new lncRNA knock-out collection in *S. pombe*. The RNAi machinery, involved in metabolism of DUTs, is conserved between humans and *S. pombe* but not in *S. cerevisiae*, making this a good model organism for studying lncRNA biology. Mutants were constructed using seamless genome editing with CRISPR/Cas9, which leaves no markers or 'scars' in the genome but is more time-consuming and difficult (Rodríguez-López et al., 2016). Mutagenesis with CRISPR/Cas9 can result in off-target effects (O'Geen et al., 2015), so every lincRNA gene was knocked out using at least two different guide RNAs. Several successful transformants for each guide RNA were included, serving as biological replicates. We set out to screen this library for phenotypes in a wide range of conditions, hoping to uncover novel context-specific functions not observed in standard conditions (Hillenmeyer et al., 2008). We expected many lincRNA mutants to show much subtler growth phenotypes than could be expected for protein-coding genes. This required an experimental approach with high throughout as well as high precision.

1.1.2 Colony screens uncover subtle phenotypes in high throughput

A first key decision when investigating growth phenotypes is whether to use liquid or solid media, where cells grow in suspension or as colonies respectively. These are two physiologically distinct states with cells in colonies showing evidence of specialisation and complex metabolic interactions (Váchová et al., 2009), accompanied with transcriptional changes (Traven et al., 2012). Zackrisson et al. (2016) have found that for salt stress and standard media a wider variation of fitness is observed on solid media compared to liquid but it is unclear to what extent this can be generalised. Whether to use liquid or solid culture for obtaining growth phenotypes is hence not a trivial matter. Both approaches additionally come with technical advantages and drawbacks. In brief, liquid cultures in 96 or 384 well plates tend to sediment, liquid evaporates without counter-measures and optical densities of cultures saturate standard plate readers (Stevenson et al., 2016). Arrayed colonies on solid media can be grown in higher densities (up to 1536 per standard plate are common) but the readout can be confounded by competition and diffusion of small molecules through the agar.

A key advantage of colony-based assays over liquid culture is that traditionally a single-timepoint (endpoint) measurement of colony area is used as an intuitive fitness proxy reflecting how well a colony has performed under the given conditions. While this means hundreds of plates can be prepared and incubated simultaneously enabling unrivalled cost-efficiency and throughput, it raises questions about the nature of the information contained in these measurements. Microbial growth is usually characterised by three key parameters, all easily obtained from liquid growth curves: length of the lag phase, maximum growth rate and final population size (also substrate conversion efficiency or yield). A true understanding for how endpoint colony sizes relate to these parameters is currently still lacking, but we have recently seen the emergence of tools that enable the investigation of growth dynamics on solid media by repeated imaging of plates: First, it has been shown that even a small number of colony size snapshots can be sufficient to fit growth models (Addinall et al., 2011; Banks et al., 2012; Shah et al., 2007). The *ScanLag* system (Levin-Reisman et al., 2014) then automated image acquisition of agar plates using commercial flatbed scanners, obtaining dense growth curves. The *Colony-live* set-up (Takeuchi et al., 2014) later improved on this by using trans-illuminating scanners (from the Epson V series) which can reveal information about colony thickness as well as footprint area. Zackrisson et al. (2016)

further improved image acquisition and data analysis with the *scan-o-matic* platform.

Colony fitness screens have been the method of choice for investigating growth phenotypes of large strain libraries in high throughput, and remarkable insights are still being obtained using this classic technique (eg Nichols et al. (2011); Costanzo et al. (2016); Brochado et al. (2018); Kuzmin et al. (2018)). In their technical implementation, a key step is the automated extraction of quantitative properties of colonies from image data. The growth curve-based approaches mentioned above have integrated this functionality into their control software but there are also a number of dedicated tools, each with their individual strengths and weaknesses. Common to all is a relatively narrow functionality with regards to the types of images can be analysed (eg acquired with cameras or scanners, in colour or in grayscale, images of colonies or more diffuse spots). *Gitter* (Wagih and Parts, 2014) and *spotsizer* (Bischof et al., 2016) determine colony sizes, while *Colonyzer* (Lawless et al., 2010) specialises in images of spots and *Iris* (Kritikos et al., 2017) quantifies morphological features relevant to biofilm formation.

Colony size data is known to be highly prone to technical variation, both within and across plates (Blomberg, 2011). Unevenness of the agar surface as well as temperature and moisture gradients result in variation within plates, while other factors such as media composition, plate handling and other experimental procedures affect results across plates. To correct for these effects, dedicated normalisation procedures have been proposed for colony fitness data. Row/column median normalisation (Baryshnikova et al., 2010) can neutralise edge effects and similarly distributed errors but assumes/requires that median fitness of each row/column is a reliable estimate of the null effect. A much more powerful normalisation strategy that does not make any assumptions and that can correct in theory any spatial effects is normalisation with a reference grid (Zackrisson et al., 2016). Here, a large number of colonies of the same control strain, placed evenly throughout the plate, are used to interpolate expected fitness for all positions. The observed fitness is then compared to the expected, resulting in an intuitive, relative fitness estimate. If the same control strain is used across plates and experiments, data from different plates become directly comparable.

All methods described so far investigate population growth, based on suitable proxy readouts such as OD or colony geometry. Viability (the fraction of live cells) is another important property of colonies which is not necessarily linked to biomass. The dead-cell stain phloxine B has for a long time been used to visualise dead yeast cells by microscopy

(Tsukada and Ohsumi, 1993) or in colonies (Matynia et al., 1998). Recently, this concept has been applied in a screening context, where colonies are assigned a 'redness' score used to detect strain-specific responses to specific conditions (Lie et al., 2018). This appears to be a very promising strategy with the potential to reveal a new dimension of colony-associated phenotypes, but little is known about the biology underlying these redness scores and very few tools are currently available to the community.

1.1.3 New tools for microbial phenomics are needed

After reviewing the available laboratory methods and computational tools, we decided that a colony-based assay of biomass and viability would be best suited to characterise our library of lincRNAs mutants. However, initially hoping to use or adapt a published pipeline, I quickly noticed that despite the long-standing popularity of colony screens and the number of publications on the topic, the tools available to researchers are actually very limited, both in their number and functionality. For example, *gitter* (Wagih and Parts, 2014) and *grofit* (Kahm et al., 2010), two popular R packages for image quantification and growth curve analysis are now out-dated and archived and no longer installed by the R package manager. Other implementations, while highly sophisticated, suffer from poor reliability, complicated/impossible installation procedures and/or lack of cross-platform compatibility. Especially the growth curve-based approaches have a very narrow scope and were designed for a specific scanner which is now out of production. These tools are also usually 'black boxes' which give the user little insight into intermediate steps performed by the software. Therefore, in our lab and presumably in the wider community, there is an urgent need for modern, streamlined tools that are easy to install and use, that work 'out-of-the-box', and that integrate into existing workflows and experimental designs. Additionally, a unified approach to analysing colony size endpoints, colony growth curves and colony redness scores would be highly desirable and make these data more comparable.

To this end, I developed the *pyphe* (python tools for phenotype analysis) toolbox which comprises of 6 command-line programs, each performing a different workflow step. The following section describes the development of *pyphe* in detail and illustrates its capabilities using a range of test datasets. This section also explores ideal strategies for phenotyping and the nature of the information contained in solid growth curves, colony size endpoints and colony redness scores. The *pyphe* pipeline is then applied to the lincRNA knock-out library in a systematic survey of condition-specific lincRNA-associated phenotypes.

1.2 Methodology

1.2.1 Pyphe provides tools for fast and precise analysis pipelines

1.2.1.1 Design principles

Based on the goals outlined in the Introduction, I set out to develop our own workflow for the analysis of colony screen data. Work on this project took place over almost four years, and initially focused on low level functions contained in a python package. When colleagues asked to run their own analyses, this approach showed its limitations. The code had to be re-structured to make it robust to bad data and user input and it required a high-level interface that could be used without programming skills. The following paragraph outlines the decisions that guided the development of *pyphe*, which overall followed modern best practice recommendations for bioinformatics software (Jiménez et al., 2017; Georgeson et al., 2019) wherever feasible and applicable.

The first key design decision was to break up the analysis pipeline into distinct steps:

1. Image acquisition
2. Image quantification
3. Analysis
4. Data Interpretation

Each step is performed by a separate computer program, with a limited, intuitively and clearly defined scope (Fig. 1.1). This limits the complexity of individual procedures and crucially means that steps of the pipeline are interchangeable with other tools. This will facilitate the uptake of *pyphe*, as researcher may decide to use the entire pipeline or incorporate selected tools into their existing workflow. Simultaneously, it future-proofs *pyphe*, eg by setting it up for compatibility with yet unpublished image analysis tools. *Pyphe* has a unique, broad scope, covering all steps of a typical experiment from scanning the plates to obtaining final summary statistics and p-values. Furthermore, it was designed to be as flexible as possible, working with different fitness proxy inputs: colony areas, redness scores and maximum slopes of growth curves (Fig. 1.1).

The second key design decision was to provide access to tools via a command line interface (CLI). *Pyphe* tools use the python *argparse* module to parse user commands, which includes powerful tools for checking their validity and for providing inbuilt documentation. Most parameters come with default values, allowing them to be invisible to non-expert users. Advanced users may decide to alter default parameters, which allows for flexibil-

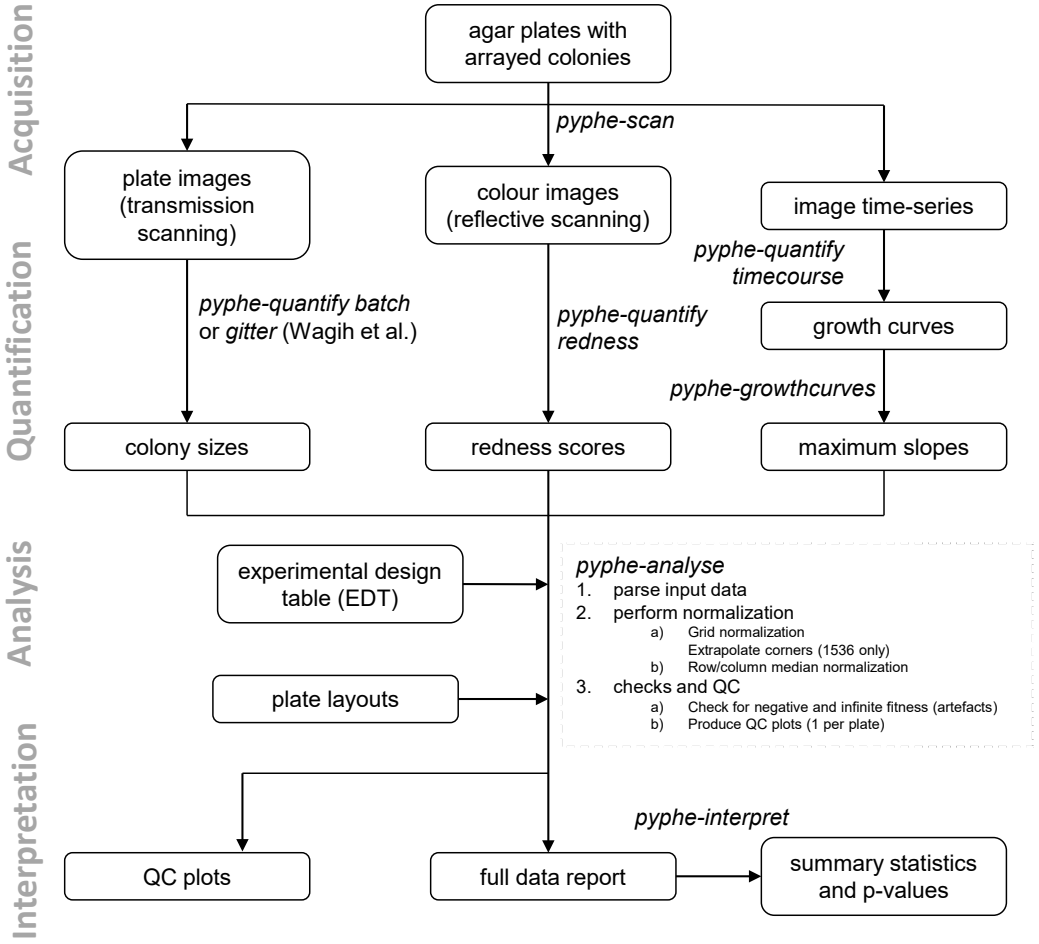


Fig. 1.1: Pyphe enables flexible and powerful data analysis pipelines. Each of the four key steps (left, vertical text) is performed by a different *pyphe* tool. Files are represented by boxes which are linked by workflow tasks (arrows) performed by different pyphe tools (in italics). The branched structure reflects the three types of input data that can be processed with *pyphe*. This modularity also allows for the integration of other tools, such as *gitter* (Wagih and Parts, 2014) for colony area measurements. *Pyphe* has a unique broad scope, supporting all steps of the analysis, from scanning to the final statistical output. Users have access to all intermediate steps in accessible, useful formats. *Pyphe* was designed with high-throughput applications in mind: It can analyse experiments containing hundreds or thousands of plates and is fully scriptable.
This figure was reproduced from Kamrad et al. (2020b) which is published under a CC-BY license.

ity while not sacrificing simplicity. It is my opinion that this approach is in this case more suitable than a graphical user interface (GUI). GUIs often struggle with cross-platform compatibility, especially when they are not browser-based, and are time consuming to build and maintain. GUIs show their strength only when interactive components are presented to the user, which is not required for any of the *pyphe* tools. Despite initially seeming more user-friendly, over-loaded GUIs with numerous tabs, menus and sub-menus are actually harmful and confusing. Crucially, the use of CLI means that *pyphe* analysis pipelines are scriptable, ie can be reproduced exactly with the same parameters and can be re-run very quickly when the input data has changed.

The third and final design decision concerns data formats. Many software packages define their own data formats and use non-human-readable (eg binary) formats to store intermediate steps. For example, *scan-o-matic* (Zackrisson et al., 2016) saves image quantification data as .npy files (which is computationally more efficient but requires users to be competent in python to read the data). *Gitter* (Wagih and Parts, 2014) produces a variation of a tab-separated table, a .dat file, which contains additional comment lines making it harder to parse and process. The full power of collaborative, open source research software can only be harnessed if a common standard for data formats is established. For example, in the field of genomics, a huge variety of innovative tools exist, each able to perform a clearly defined, narrow task, because the field agreed on the data formats they operate on (ie fastq, bam/sam and vcf files). *Pyphe* uses human-readable, standard comma-separated tables (csv) throughout and adheres to data-science best practices. Specifically, it uses tidy formats (Wickham et al., 2014) wherever possible which makes the data ideally suited for statistical analysis and plotting with packages such as seaborn and pandas. Giving the user direct access to all data produced along the pipeline allows users to check how individual steps and to re-use and further analyse data in other ways. Based on these design principles, six *pyphe* tools were developed which together support all major steps of colony screen analysis workflows. The following sections describe their core concepts and algorithms and demonstrate their capabilities.

1.2.1.2 Image acquisition with pyphe-scan

In the first step of the *pyphe* pipeline, images of plates and colonies are acquired. Standard digital cameras can and have been used for this but the image will depend strongly on the lighting of the object, which will in turn trigger camera-internal adjustments to focus,

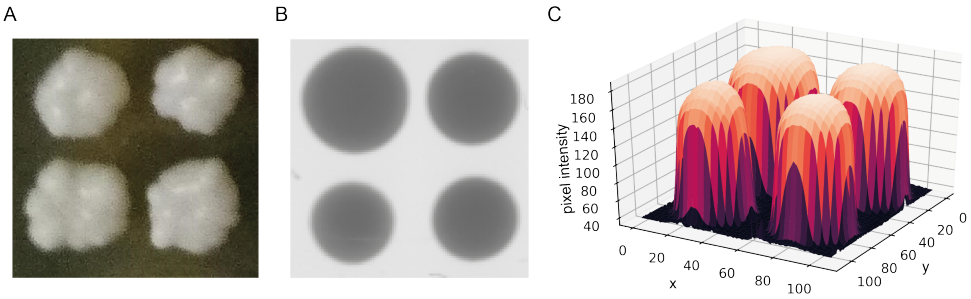


Fig. 1.2: Transmission scanning produces high quality images. **(A)** Picture of colonies growing on agar taken with a phone camera. Reflections off the colony and agar surface results in colour artefacts. **(B)** Picture of colonies growing on agar taken by transmission scanning. Colonies appear dark against a light background. The original plate carried 1536 colonies of *S. pombe* wild strains grown for two days on rich YES media. The image was scanned in 8-bit grayscale and at resolution of 600dpi. **(C)** Inverted pixel intensity values across the image from (B). Pixel intensities from transmission scans can be used to estimate colony thickness.

brightness and white balance. It is therefore hard to achieve consistent image properties and quality, especially for large experiments conducted over extended timespans. The use of gel-doc systems can alleviate some of these problems, but only to an extent. We, and others (Takeuchi et al., 2014; Zackrisson et al., 2016), have found that transmission scanning with commercial flatbed scanners is ideally suited for this task. With transmission scanning, the light source and camera/sensor are on opposite sides of the object. The lighting from the scanner lid produces even and consistent light against which colonies appear much darker than the background. It does not produce shadows, or reflective artefacts on the surface of colonies (Fig. 1.2). The darkness of a pixel reflects the thickness of the colony at that position, which has been used to approximate colony mass (Takeuchi et al., 2014) and even cell number (Zackrisson et al., 2016).

Maintaining a constant relative position of the plate and colonies within the image frame is a key challenge to solve during image acquisition. Image analysis will typically fail if plates are rotated relative to the image frame and the presence of plate borders will impact image analysis with many tools. For experiments containing hundreds or thousands of plates manual cropping and rotating of images is not an option, so it is crucial to ensure that the position of plates in the image frame is consistent so that this step can be automated. I have implemented a similar solution to those used by Zackrisson et al. (2016), where a custom-made device is laser-cut from acrylic glass and used to hold plates in place. The design of this ‘fixture’, which accommodates up to 4 standard plates per scan, is published

on the *pyphe* GitHub repository.

Pyphe-scan acquires images using the Linux Scanner Access Now Easy (SANE) interface, similarly to Zackrisson et al. (2016). *Pyphe* in theory works with every scanner supported by SANE (which are almost all commercially available scanners) but has been developed an tested with the Epson V800, the newer model in the series used previously by others (Takeuchi et al., 2014; Zackrisson et al., 2016). Apart from this similarity, *pyphe-scan* is a completely new implementation. *Pyphe-scan* facilitates imaging stack of plates of arbitrary size, by prompting the user to switch plates between scans, and will produce pictures of individually plates which are numbered consecutively. A number of parameters, eg the scanning resolution, image output format and the number of plates to be scanned are set by the user. *Pyphe-scan-timecourse* is a small variation of *pyphe-scan* which supports timecourse imaging, where images are acquired from the same scanner at a fixed time interval. Following *pyphe's* design principles, users have access to all intermediary data processing steps. *Pyphe-scan* crops and rotates individual plate images from the raw scan (which usually contains four plates). Plates are placed in the scanner with the colonies facing up but the image is taken from below, so plates need to be flopped (mirrored along the vertical midline). 'Under the hood', these steps are performed by *ImageMagick*, a powerful command line tool for image manipulation.

1.2.1.3 Image quantification with pyphe-quantify

In the next step of the *pyphe* pipeline, quantitative characteristics of individual colonies are extracted from images. *Pyphe* contains a versatile and fast tool for image quantification built on the powerful *scikit-image* library (van der Walt et al., 2014). Depending on the choice of fitness proxy, *pyphe-quantify* extracts colony area, colony redness or the sum of pixel intensities over each colony. In timecourse mode, the last image/timepoint is used to reconstruct growth curves from image timeseries. *Pyphe-quantify* by default analyses all jpg images in the current directory and can process thousands of images per hour. The following paragraphs describe the image analysis procedures in detail.

Detection of colonies against background by thresholding. In images acquired with transmission scanning, colonies appear dark against a light background. They can easily detected by applying a binary threshold to the image, which is determined dynamically using the Otsu method (Otsu, 1979), followed by the identification of connected components. Analysis of images acquired in reflective colour mode for colony redness analysis require a

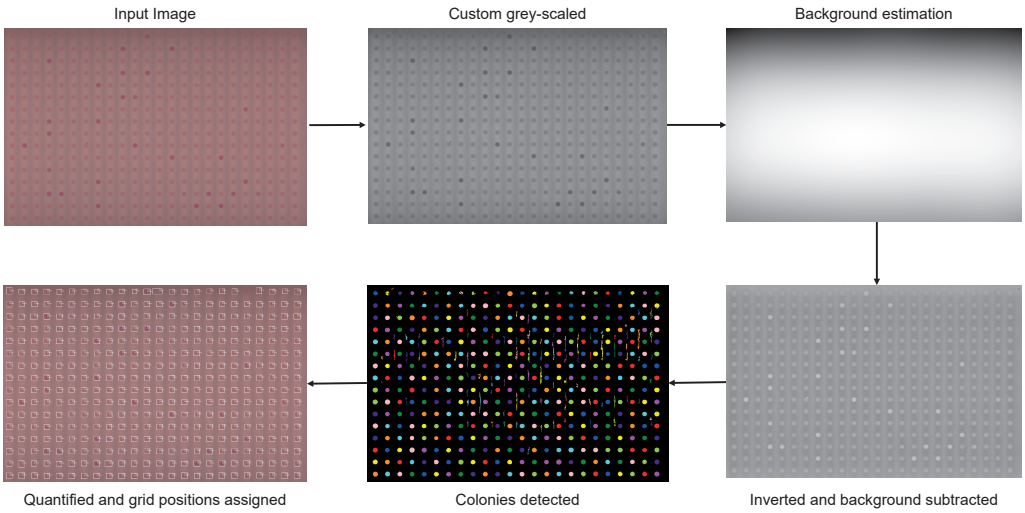


Fig. 1.3: Quantification of colony redness requires refined image analysis. Images scanned in reflective colour mode for colony redness analysis exhibit faint colonies and uneven lighting. Images first transformed to grayscale by multiplying the red, green and blue channels by 0, 0.5 and 1 respectively and taking the sum. This makes colonies appear stronger against the background. Next, uneven brightness across the image is corrected by subtracting the image background, which estimated by strongly blurring the image with a Gaussian filter. Colonies are detected by local thresholding and the mean pixel intensity within each colony is reported. *Pyphe-quantify* produces a QC image in which detected colonies and their score are marked. This figure was reproduced from Kamrad et al. (2020b) which is published under a CC-BY license.

refined approach as colonies are much fainter and images contain uneven lighting (Fig. 1.3). Images are first transformed to grayscale by multiplying the red, green and blue channels by 0, 0.5 and 1 respectively and taking the sum. Redness images suffer from uneven brightness across the image, which is corrected by background subtraction. For redness images, *pyphe-quantify* uses local, adaptive thresholding to increase the sensitivity of colony detection.

Assignment of colonies to grid positions. Once potential colonies have been identified, they need to be matched to grid indices (colony row/column indices). *Pyphe-quantify* can either use user input to determine the expected position of colonies (the grid positions) or guess them automatically based on the image data. The manual method is more robust and is preferred when the image has many missing colonies. The user needs to enter the number of rows, columns and the coordinates of the top-left and bottom-right colony in the image. This means that colony positions in the image stack to be analysed need to be more or less identical and this is the case if *pyphe-scan* is correctly configured. Alternatively, grid position can be determined automatically by identifying peaks across pixel rows and

columns across the image. Peaks are identified using the find_peaks function in *scipy* (Virtanen et al., 2020) and the inter-peak distance is determined by taking the trimmed mean of all peak distances. A cosine function with the same periodicity is generated and slided over all possible positions in the image and the global best fit is used to compute grid positions. Next, a distance matrix between putative colonies identified in the image and the grid positions is computed, and each grid position is matched to the closest colony within a distance threshold around it.

Data collection and output. Multiple parameters are then extracted for each colony. Depending on the running mode, this includes the main output -colony area, mean colony redness or sum of pixel intensities over time- as well as additional readouts such as the assigned grid position, colony position in the image, colony circularity, colony mean pixel intensity and others. All data is written to a standard csv file. In the batch modes, a single file per plate is created which allows the user to easily check and re-use data from individual files. *Pyphe-quantify* also produces QC images, marking the position, size and grid assignment of any identified colonies. In redness mode, the assigned redness score is reported.

Pyphe obtains similar colony areas as *gitter*. In order to benchmark the performance of *pyphe-quantify* against an established image analysis tool (Wagih and Parts, 2014), a set of test images was obtained by scanning a single plate in grayscale transmission scanning mode in 20 minute intervals for 48hrs. The plate carried 57 diverse *S. pombe* wild strains growing in 1536 format (approximately 16 replicates per strain). All images were analysed with *gitter* and with *pyphe-quantify* in batch mode. The results obtained (across all 144 images) are very tightly correlated and can be considered identical (r=0.9991, Fig. 1.4A). However, it is important to note that absolute values are not identical as *pyphe* is in general a bit more conservative during thresholding which results in the outermost, faintest edges of colonies not counted towards the colony area. In practice, this is of no relevance in almost all application, as the dependence is linear which results in similar relative fitness estimates after normalisation.

Redness scores reflect the fraction of viable cells in colonies. The ability to extract colony redness scores from scan images is a distinguishing feature of *pyphe-quantify*. Phloxine B is known to stain dead cells (Matynia et al., 1998), but how are redness scores and colony viabilities quantitatively related? To investigate this, the composition of 23 colonies of varying redness was analysed by imaging flow cytometry. Colonies were picked from a

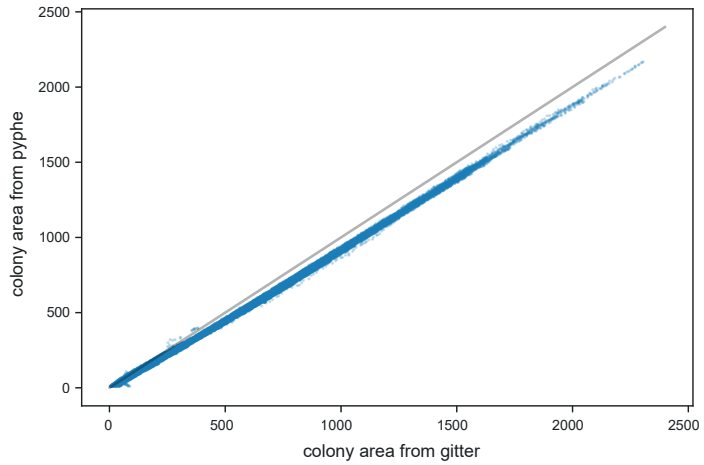


Fig. 1.4: Pyphe-quantify produces colony area measurements similar to *gitter*. *Pyphe* colony area measurements are tightly correlated with those obtained with *gitter* (r=0.9991, Pearson) across 144 images with approximately 1536 colonies each, demonstrating the robustness and reliability of colony area measurements obtained with *pyphe-quantify*. While the two readouts are linearly correlated, colony areas with *pyphe* are smaller than those reported by *gitter* (due to stricter thresholding). This makes no difference in practice once normalisation has been applied. This figure was reproduced from Kamrad et al. (2020b) which is published under a CC-BY license.

plate containing 238 S. pombe knock-out mutants grown for two days on rich media with 5mg/L phloxine B. The analysis distinguished three populations based on their phloxine B staining intensity (Fig. 1.5A+B). The population of cells with the lowest level of staining could be identified (based on the simultaneously acquired brightfield image) as visibly damaged, lysed cells or cell debris. Viable cells, the largest population, showed a medium, background level of staining while dead cells stained strongly. The fraction of live cells in a colony was found to anti-correlate well with colony redness scores (r=-0.88, Fig. 1.5C). For this analysis, redness scores were normalised using the approach outlined in the following section. In some colonies, specifically those with a lower fraction of live cells overall, lysed cells make up the majority of non-viable cells. The tight anti-correlation between the fraction of live cells and colony redness score suggest that lysed cells, while unstained in the flow cytometer, do contribute to colony redness. This is most likely due to the washing step applied before flow cytometry analysis, which washes out any dye, which does not happen in the colony. This is supported by the observation that the fraction of live and lysed cells together correlates less strongly with colony redness (r=-0.78, Fig. 1.5D). The dynamics and physiological state allowing phloxine B to enter the cell are not entirely understood. Therefore, cell populations from colonies based on phloxine B staining were compared to those

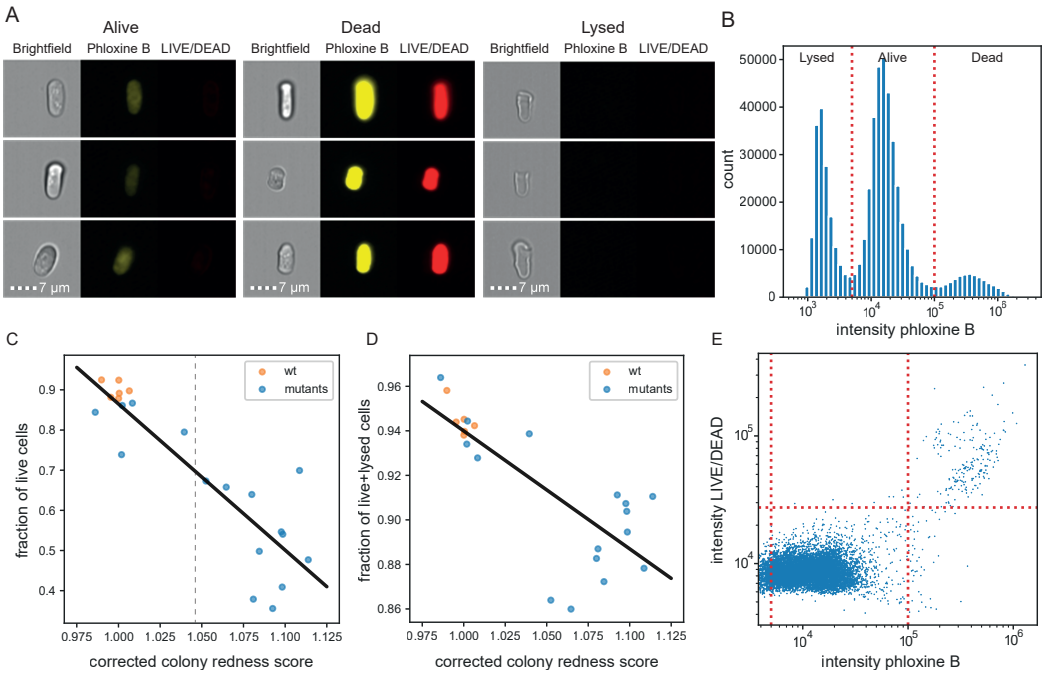


Fig. 1.5: Redness scores accurately reflect colony viability. **(A)** Exemplary data from imaging flow cytometry of resuspended colonies illustrating the three distinct cell populations: alive, dead and lysed. **(B)** These three populations are clearly distinguishable from phloxine B staining. Counts summarise 541,432 individual cells from 23 colonies. **(C)** The fraction of live cells per colony is strongly anti-correlated with normalised colony redness score for 23 colonies of different *S. pombe* knock-out mutants. Some grouping of colonies (dashed line) is apparent, with a lower correlation in each subgroup than overall (r=-0.41 and r=-0.33 for left and right group respectively, Pearson correlation). **(D)** While lysed cells are unstained in the flow cytometer, due to the preceding washing step, they do contribute to colony redness. Accordingly, the correlation of live and lysed cells with colony redness score is lower than that for live cells only (C). The correlation specifically is worse for colonies with a high redness score, which contain larger fractions of lysed cells. **(E)** Phloxine B is in close agreement (accuracy=99.3%) with the chemically distinct commercial LIVE/DEAD stain. Data is based on 9582 cells from a single wild type colony.
Panels of this figure were reproduced from Kamrad et al. (2020b) which is published under a CC-BY license.

obtained with a chemical distinct, established commercial dead cell stain (LIVE/DEAD, Thermo Fisher Scientific). For colonies of the standard lab strain and disregarding lysed cells, both dyes were in near-perfect agreement (accuracy=99.3% using LIVE/DEAD classification as ground truth, Fig. 1.5E). Overall, colony redness scores provide an accurate readout reflecting the number of live cells in a colony.

Images can be compressed without affecting results. *Pyphe* supports multiple image formats. In particular, users have the choice between the lossless image format tiff, which preserved the original scanner data exactly, or the 'lossy' jpg format, conversion to which can result in data loss due to compression. Using the standard conversion parameters of

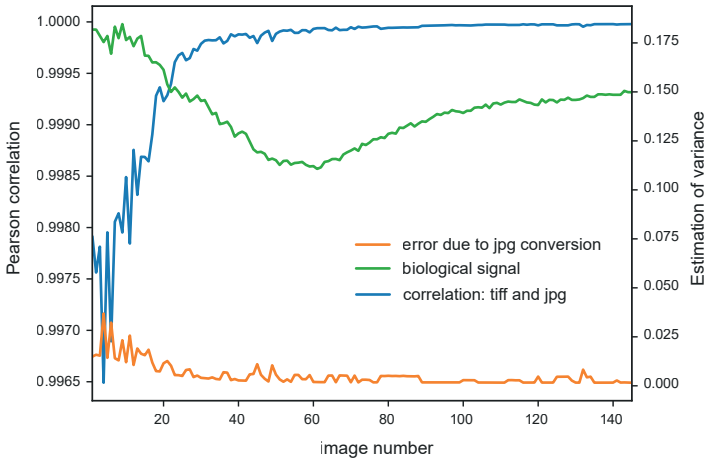


Fig. 1.6: Image compression does not affect fitness readouts. 144 original images of 1536 colonies in lossless tiff format were compressed to jpg format, resulting in a significant 20-fold reduction in disk space requirements. Colony areas in both image sets were analysed with *pyphe-quantify* and compared. Results are in very close agreement (blue line, global Pearson correlation of 0.999964) and any errors introduced by conversion (orange) are negligible compared to the biological variation of colony areas (green).
This figure was reproduced from a rebuttal letter for Kamrad et al. (2020b) which is published under a CC-BY license.

ImageMagick, the size of a single image of a plate shrinks from approximately 4MB down to 180-580KB during conversion. This approximately 20-fold reduction is storage space requirements is significant, especially for larger experiments containing thousands of plates. The set of test images was analysed in jpg and tiff format and the obtained colony sizes were compared. Across all images, the Pearson correlation was 0.999964 and the median relative error

$$\mathrm{re} = \frac{\mathrm{abs}(\mathrm{size}(\mathrm{jpg}) - \mathrm{size}(\mathrm{tiff}))}{\mathrm{size}(\mathrm{tiff})}$$

was 0.00333. The results were consistent across the timeseries (Fig. 1.6), even for early images which are harder to analyse due to smaller, fainter colonies. These detected deviations need to be seen in relation, and turns out to be negligible compared to the biological signal, estimated by the median absolute deviation of all colonies in each plate.

1.2.1.4 Growthcurve analysis with pyphe-growthcurves

For the downstream analysis of growth curves extracted from image timeseries with *pyphe-quantify*, a simple quantitative fitness readout needs to be extracted. For this, *pyphe* contains a simple tool for extracting maximum slopes and lag phases: *pyphe-growthcurves*. Growth curves consist of the sum of background-subtracted pixel intensities within the colony for each timepoint. This is closely linked to colony biomass but not directly linearly related

to cell number. For this reason, *pyphe-quantify* does not fit parametric model of microbial growth to the observed data. The maximum slope is extracted using a simple sliding window approach, whose width can be set by the user. For each window, a line is fitted to the growth data and the overall highest scoring window is reported. A simple threshold-based method for determining lag phases is implemented, reporting the timepoint at which each colony crosses an absolute or a relative value (relative to the initial biomass). *Pyphe-growthcurves* produces a csv table listing extracted growth curve parameters for each colony which can be directly used for further analysis in *pyphe*. *Pyphe-growthcurves* uses a simple input format and has also been successfully used on growth curves obtained from plate reader measurements of liquid cultures in multi-well plates.

1.2.1.5 Data analysis with pyphe-analyse

The next step in the *pyphe* pipeline performs two key functions. It normalises data within and across plates to make them directly comparable and it aggregates data and meta-data across all plates of an experiment, producing a comprehensive data report. *Pyphe-analyse* contains parsers for different fitness inputs (colony area endpoints obtained by *gitter*, colony areas and redness scores obtained by *pyphe-quantify* and maximum slopes of growth curves obtained with *pyphe-growthcurves*), making it a powerful and versatile tools for different fitness screen scenarios.

Control grids enable efficient normalisation with minimal underlying assumptions. The most common type of technical within-plate variation are edge effects. Here, colonies on the outer edge of the grid grow better as they experience decreased competition over nutrients from neighbouring colonies (Fig. 1.7A). This error follows a clear row/column pattern and can be corrected by dividing each colony fitness readout by the median of its row and column (Baryshnikova et al., 2010). This normalisation requires that the null-effect of each row/column can be estimated reliably, which requires a sufficient number of colonies (the 96 format is unsuitable in our hands) and requires that most colonies in every row and column show no growth effect. This is usually the case for work with knock-out libraries, where the vast majority of mutants behaves like wild type in a given condition, as long as strains with abnormal growth are randomly distributed in the assay plate and not concentrated in any row/column. Besides these limitations regarding the strain library and its arrangement, row/column-wise correction cannot correct for more complex patterns of technical noise. Colonies will usually show substantial variation in fitness proxies across

a plate, due to technical factors such as temperature and moisture gradients or unevenness of the agar surface (Zackrisson et al., 2016; Kamrad et al., 2020b). The fact that an evenly spaced grid of control strains can approximate the null effect in different positions of the plates is the basis for the spatial grid normalisation first proposed by Zackrisson et al. (2016). Based on the control grid, the expected colony size of a control strain in each condition is interpolated using a curvature-minimising cubic spline fit (Fig. 1.7B). The value of this reference surface in each position is then compared to the actually observed colony size. This results in a relative fitness readout, where 1.0 signifies a similar fitness as the control strain (Fig. 1.7C). This strategy is highly effective in reducing noise and is suitable to diverse strain layouts (Fig. 1.7D, Zackrisson et al. (2016); Kamrad et al. (2020b)).

Pyphe improves data completeness during grid correction. While the core functionality of the grid correction implemented in *pyphe* is identical to that of *scan-o-matic*, *pyphe* implements a key improvement that prevents 'blind spots' during normalisation. Working in 1536 format, the *scan-o-matic* workflow uses a grid of 384 control colonies. This necessarily leaves two edges of the plates which are not covered by control strains. In these positions, the value of the reference surface is undetermined, unless it is guessed, eg by extrapolating the nearest value (which cannot account for the edge effect). *Pyphe* recommends instead to place two 96 grids with opposite offsets (Fig. 1.8A). This decreases the amount of colony positions required for normalisation by half and already greatly increases the area covered by the normalisation surface. For the two remaining corners still not covered by the grid (Fig. 1.7B), I have devised a precise extrapolation method based on statistical learning. Across all plates of the experiment, a linear model is trained where the dependent variable (the label) is the observed raw size of the control strain in the corners which are covered by the reference grid and the features/predictors are the sizes of the next horizontal and vertical control strain. These models typically predict corner colony sizes with $>90\%$ accuracy for the training set and can be applied to the other two corners to predict expected control strain colony sizes there (Fig. 1.8B). These improvements significantly improve the data completeness after normalisation and mean that libraries can be used without the need for re-arrangement. Furthermore, it increases the throughput of the workflow by leaving more space for assay colonies. We usually spare an additional 96 colonies for an additional internal control to monitor technical noise levels and normalisation efficiency across plates.

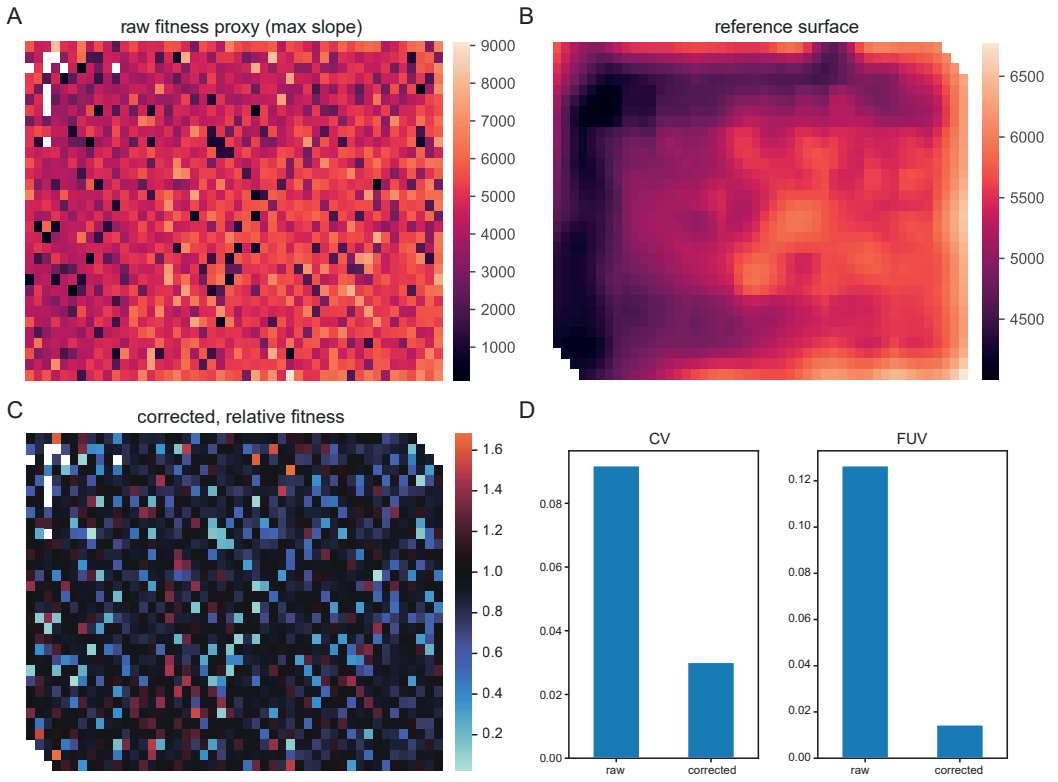


Fig. 1.7: Grid correction effectively reduces technical noise. **(A)** Maximum slopes recorded from a single plate containing 1536 colonies of *S. pombe* wild strains. The standard lab strain 972 was used as control and placed in 192 positions, evenly dispersed across the plate. **(B)** An interpolated reference surface computed based on the control strains only. The surface estimates the null effect, ie the expected maximum slope of a control strain, for every position. **(C)** Corrected, relative fitness estimates are computed by dividing the real observed maximum slope by the reference surface. **(D)** This correction procedure is very effective in reducing technical noise in the data. The coefficient of variation (CV) of 96 additional control strains, not used to construct the reference, decreased by approximately four-fold. The fraction of unexplained variance (FUV), the ratio of the variance of the control strains divided by the variance of all colonies, decreased more than six-fold.

Grid correction results in over-correction. Globally, grid correction typically reduces the noise by 4-fold (Fig. 1.7D), but it can introduce normalisation artefacts. In particular, I have observed a secondary edge effect (Fig. 1.9). Its origin lies in the fact that colonies in the second row/column are compared to a predicted value heavily influenced by colonies on the outer-most edge. If the plate layout is suitable, this can be corrected with an additional row/column normalisation, otherwise this artefact has to be tolerated as the 'lesser of two evils'. *Pyphe* is designed as a toolbox and requires the user to carefully consider plate layouts and normalisation strategies. It offers the choice of applying one or the other, or both, or neither of the normalisations.

The secondary edge effect is one example of a more fundamental risk associated with

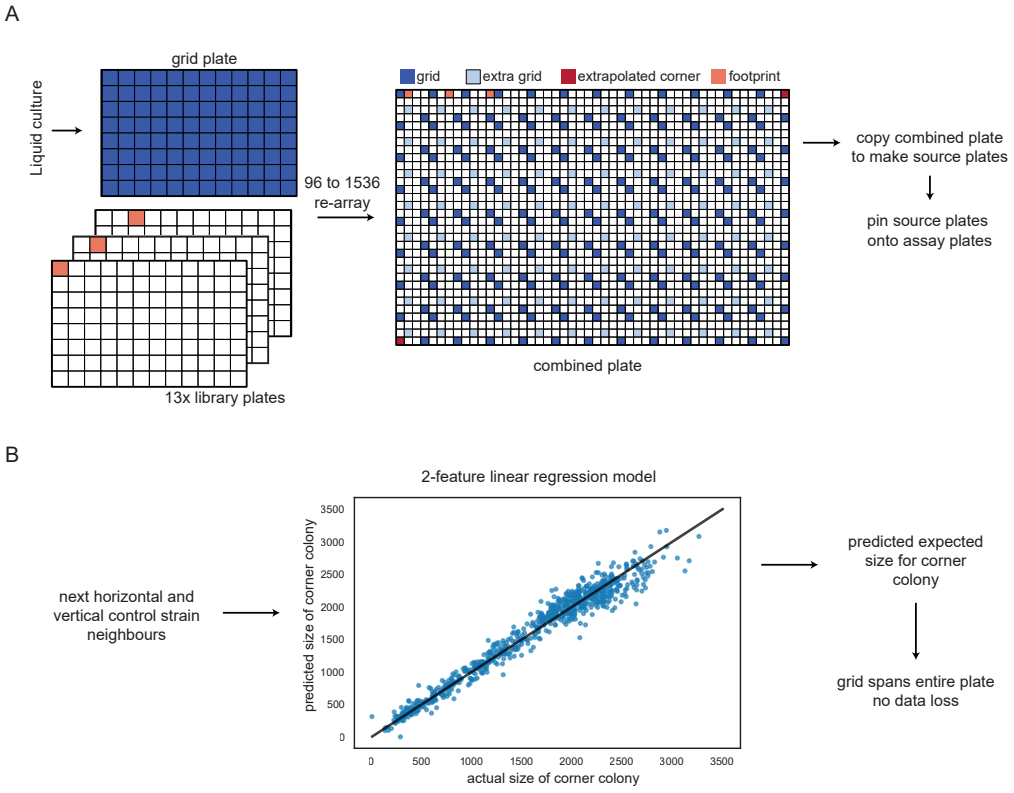


Fig. 1.8: Optimal grid placement and extrapolation. **(A)** *Pyphe* recommends placing two 96 grids of control strains with opposite offsets onto 1536 plates. This results in almost the entire plate being covered by the grid and leaves more positions for colonies to be assayed, compared to the approach taken by Zackrisson et al. (2016). An additional control grid can be placed in the plate for general quality control and noise monitoring. **(B)** The two remaining corners not covered by the grid (cf Fig. 1.7B) can be extrapolated with high accuracy based on the horizontal and vertical control neighbours. Across all plates of an experiment, *pyphe* trains a linear regression model on the other two corners, typically achieving >90% accuracy. In this case, which is based on an unpublished experiment containing hundreds of plates, the exact model was y = 1.24 * horizontal neighbour + 0.16 * vertical neighbour - 34, and the R^2 was 0.96. The symmetry of the plate allows this model to be applied to the corners without control colonies, which are then used to extrapolate and complete the reference surface.
Panel A and part of panel B are reproduced from Kamrad et al. (2020b) which is published under a CC-BY license.

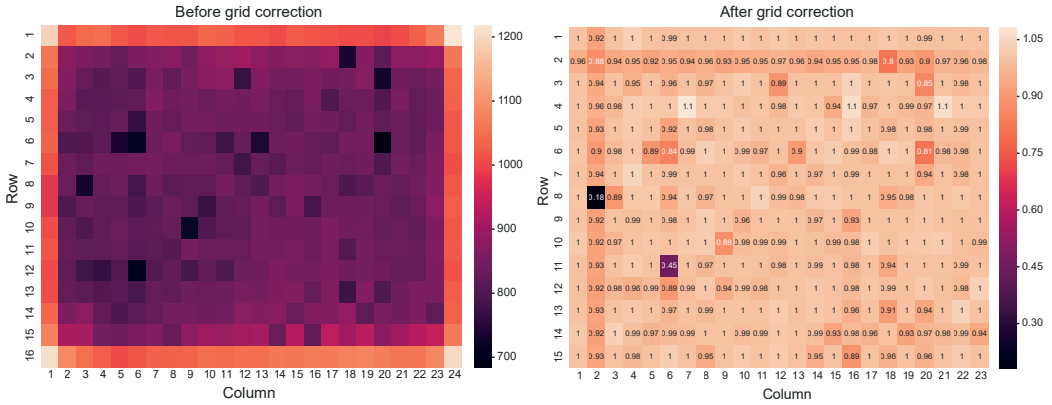


Fig. 1.9: A secondary edge effect is an artefact of grid normalisation. Before correction, plates show a strong edge effect (left). Grid normalisation removes this edge effect but introduces a weaker, secondary edge. This artefact of the correction results in lower fitness estimates of colonies in the second row/column. Plots are based on median values for each position across hundreds of plates containing *S. pombe* knock-out mutants. A single 96 control grid was placed in the top-left offset of a 384 plate, leaving the left and bottom edge not covered by the grid. The secondary edge is therefore only visible on the top and on the left.
This figure was reproduced with minor changes from Kamrad et al. (2020b) which is published under a CC-BY license.

grid correction. Because colonies grow close together on the agar surface, competition for nutrients is a large factor determining colony size. Presumably, a control grid colony growing next to a fast-growing colony will have decreased access to nutrients. When the fast-growing colony's relative fitness is then determined by comparing it to the control colony, the resulting fitness proxy will be inflated. The inverse argument is equally plausible of course. To check whether this is a problem in practice, we compared corrected colony sizes with those of its neighbours. After normalisation (grid with subsequent row/column median normalisation), no correlation between the corrected colony size and the mean if a colony's neighbours is visible (Fig. 1.10). In conclusion, *pyphe* supports various normalisation strategies, which have been shown by us and others to effectively neutralise experimental noise within plates.

Pyphe performs QC and exports data in a useful format. Both correction strategies can lead to spurious data, especially if the input data is of low quality or incomplete. Grid correction can occasionally result in corrected fitness estimates which are negative. This is because the interpolated reference surface dips below zero between very small grid colonies, due to its curvature-minimising properties. Row/column median normalisation can result in infinite values of the majority of colonies in a given row/column have size 0. *Pyphe-analyse* checks for such cases and throws warnings, prompting users to check the data input before

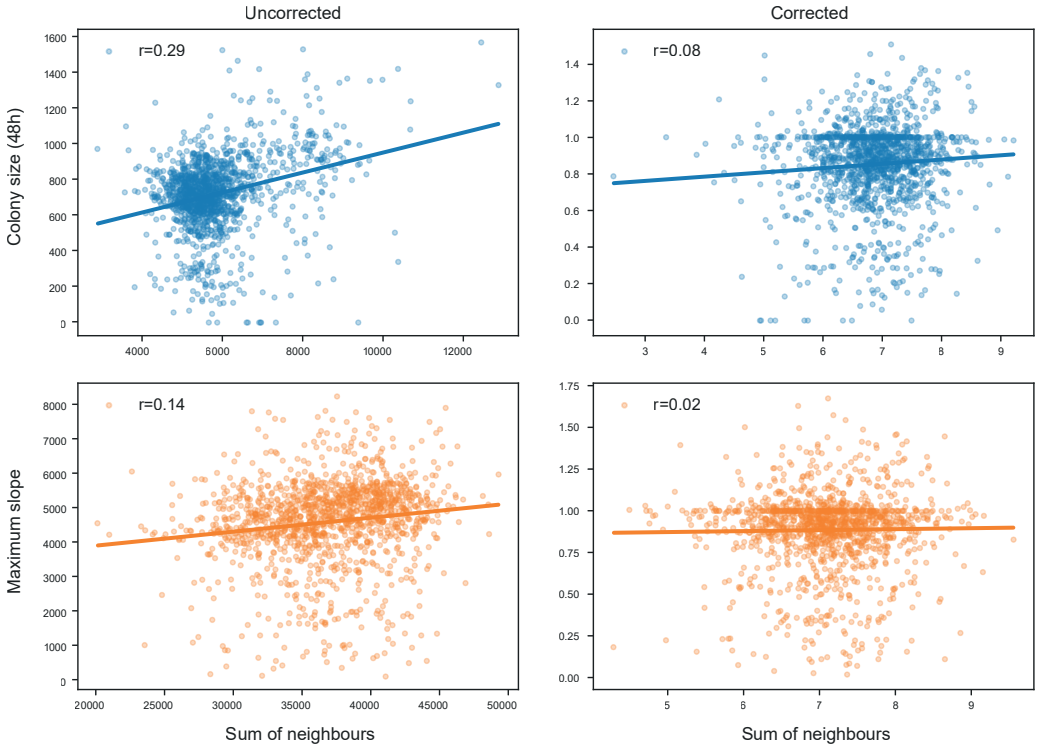


Fig. 1.10: No evidence for over-correction during normalisation. Grid normalisation compares the observed fitness of a colony to that of neighbouring controls. This comes with a risk of over-correction as colonies compete over resources. Before correction colony fitness is positively correlated to the sum of its neighbours (left). This clearly indicates that regional technical biases dominate over the competition effect (and this is one of the underlying assumptions of the grid correction). This correlation is stronger for endpoints than for maximum slopes. After correction this correlation is neutralised and crucially no negative correlation is observed. Over-correction therefore appears not to pose a significant problem in practice. Data is based on 1536 *S. pombe* wild strains grown on rich YES media for 48hrs.
This figure was reproduced with minor changes from Kamrad et al. (2020b) which is published under a CC-BY license.

setting spurious values to not applicable (NA).

Besides correcting within-plate variation, the use of relative fitness measures makes results directly comparable between plates and batches, given that the same control strain has been used. *Pyphe-analyse* was written to process groups of plates (which can easily be hundreds or thousands in size) simultaneously and can automatically incorporate layout information and meta-data into the analysis. Users provide a table listing each plate and associated information which is combined with the processed data to form a comprehensive output data report in tidy format, containing for each colony on a single line raw and normalised values, details of the normalisation and any data provided by the user in the input table. *Pyphe-analyse* assists user with quality control (QC) by producing a pdf with

heatmaps of all numeric data associated with each plate, which can be used to understand unusual or unexpected results.

1.2.1.6 Data interpretation with pyphe-interpret

Data reports generate by *pyphe-analyse* can easily contain millions of data points of numerous strains in numerous conditions. The next challenge was therefore to design an interpretation tool which is computationally efficient and sufficiently versatile to interrogate the data in the different ways, yet is reasonably straight-forward to use. *Pyphe-interpret* walks this line and produces comprehensive summary statistics and differential fitness analyses. It implements t-tests, including multiple testing correction, for differential fitness assessment and this is implemented in a fast, vectorised form. It was written to directly operate on data reports written with *pyphe-analyse* but can in theory be used with any table in tidy format. The user chooses the grouping variable, the axis variable along which to apply the test and the name of the control to test against. This allows *pyphe-interpret* to be used for answering two distinct questions: 'Which strains grow differently from the control strain (tested for all conditions individually and in parallel)?' and 'Which strains grow differently in a given condition compared to the same strain in a control condition (tested for all strain individually and in parallel)?'. The latter way of assessing differential fitness is usually preferred as it tests for condition-specific growth effects, effectively normalising out basal growth differences between strains which are already apparent in the control condition. *Pyphe-interpret* produces a comprehensive output report listing summary statistics (median and mean fitness, observation count, standard deviation) as well as p-values and effect sizes associated with t-tests.

Pyphe supports data analysis workflows from start to end, ie from scanning to differential fitness statistics. It is up to users to assemble pipelines from the tools available within (and potentially outside of) *pyphe*, which provides flexibility while simultaneously limiting the complexity of each individual tool. *Pyphe* was written specifically with very large experiments in mind, it is computationally efficient and fully scriptable/reproducible. *Pyphe* supports growth curves, as well as endpoint colony sizes and colony redness scores as input. It is the first and only solution with such a broad scope and range of functionalities.

1.2.2 Endpoints and maximal slopes of growth curves provide similar information

Pyphe was designed to be used in a wide range of phenotyping scenarios, using different fitness proxies as input. Specifically, it supports the assessment of biomass formation by either maximum slopes of growth curves or by endpoint colony areas. Previous, solely growth curve-based methods (Takeuchi et al., 2014; Zackrisson et al., 2016) have claimed the superiority of this approach, primarily based on the observation that maximum slopes contain lower levels of technical variation. However, obtaining growth curves is in many aspects more costly and time-consuming. A large number of scanners is required to conduct even medium sized screens in a reasonable time and these needs to be placed in temperature-controlled environments. Most larger experiments will have to be divided up into many batches, with substantial work involved in preparing fresh library plates for each batch and increasing the likelihood of batch effects. Growth curves produce large number of images and have a data footprint that is approximately 200 times larger (assuming imaging every 20 minutes for three days). For endpoint measurements supported by *pyphe*, hundreds of plates can be scanned and processed on a single day with a single scanner. This allows for many more replicates to be included in the analysis, thereby potentially compensating for the increased noise associated with endpoints.

In order to investigate and compare the accuracy of *pyphe*-enabled growth curves and endpoint measurements, 57 genotypically and phenotypically diverse *S. pombe* wild strains (Jeffares et al., 2015) were grown in approximately 16 biological replicates in 1536 format in 8 different conditions. These were 4 different carbon source mixes and 4 different nitrogen sources, chosen specifically, like the strains, to produce as diverse growth dynamics as possible. Each timepoint was analysed as an individual 'endpoint' and the maximum slope of the growth curve was extracted using *pyphe-growthcurves*. Noise indicators (CV and FUV) of uncorrected endpoints depend strongly on the timepoint and are lowest towards the end of the rapid growth phase, as illustrated by data from standard YES media (Fig. 1.11A). As growth progresses slowly, competition effects and edge effects increase the level of technical noise. Grid normalisation efficiently corrects the effects, largely maintaining low noise levels until the end of the experiment. Noise levels of grid corrected maximum slopes were lower than those of endpoints throughout (Fig. 1.11A). This trend held true considering all 8 conditions, which puts *pyphe* in agreement with previous methods (Fig. 1.11B) (Takeuchi

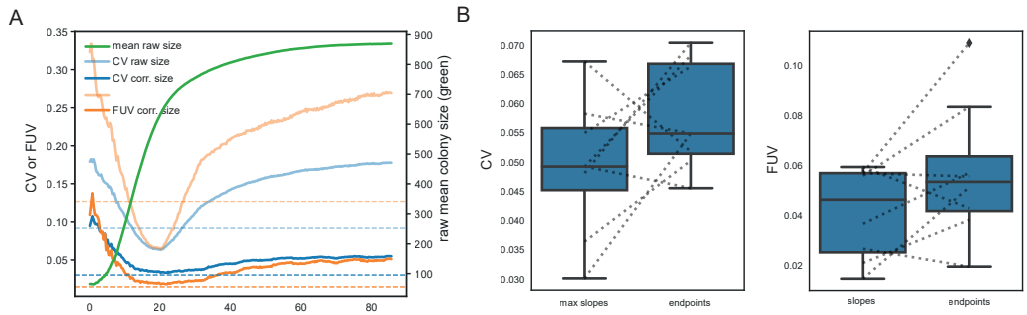
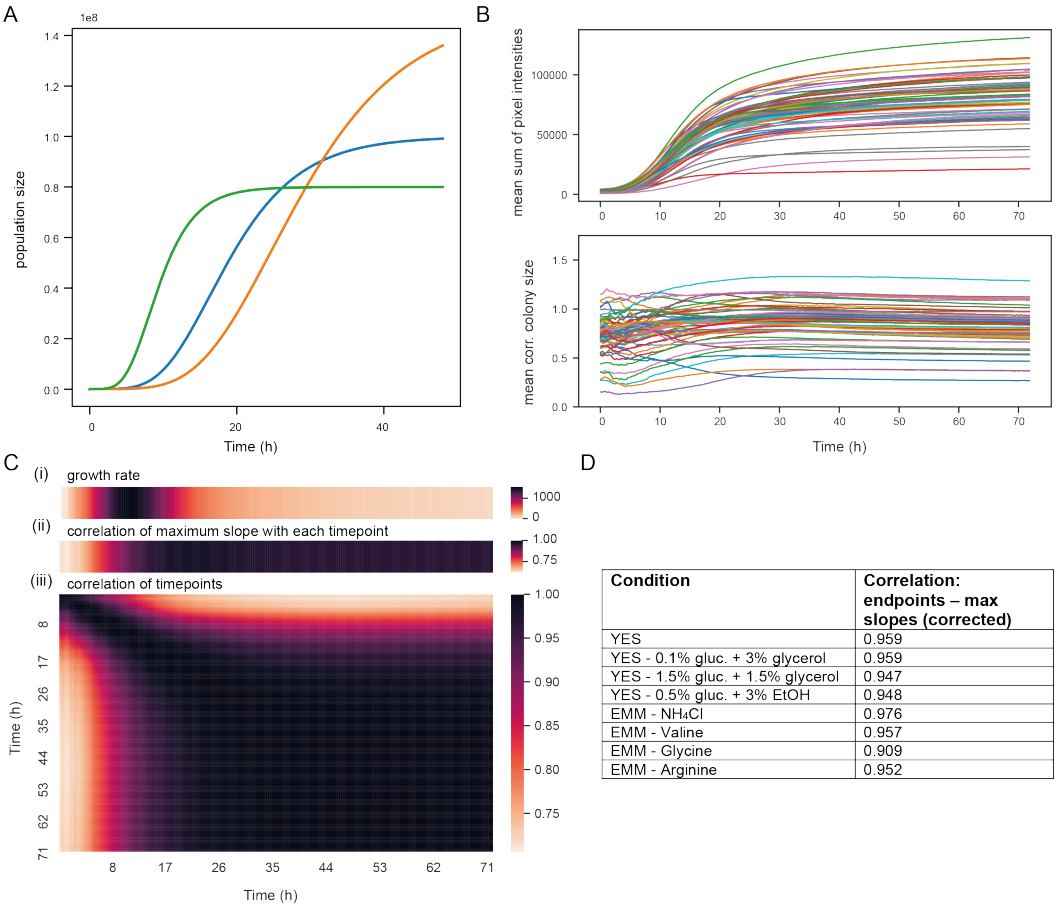


Fig. 1.11: Comparison of noise levels in endpoints and growth curves. **(A)** Technical CVs and FUVs of each individual timepoint, before and after grid correction (solid lines). The CV and FUV of grid-corrected maximum slopes are shown as dashed lines. Grid correction strongly reduces noise in the data and maintains noise near its global minimum for later timepoints. **(B)** Boxplots comparing CVs and FUVs of maximum slopes and endpoints across eight conditions. Dashed lines indicate points belonging to the same condition. These plots indicate that maximum slopes of growth curves usually have lower technical nose than endpoints. All boxplots in this thesis show the median as the central line, the quartile range as the box and the rest of the distribution as whiskers. Points more than 1.5 times the inter-quartile away from the high and low quartiles are considered outliers and are shown individually.
Both panels were reproduced from Kamrad et al. (2020b) which is published under a CC-BY license.

et al., 2014; Zackrisson et al., 2016).

The second argument put forward in favour of growth curves is based on based on the theoretical consideration that endpoint results could depend on the incubation time for strains showing different growth dynamics (Fig. 1.11A). The 57 wild strains are genetically diverse and have been isolated from various niches (Jeffares et al., 2015) and therefore ideally suited to evaluate the if this poses a problem in practice. Real growth curves on standard YES media showed little evidence of growth curves 'crossing over' after approximately 20h when rapid growth had stopped (Fig. 1.12B). A correlation analysis of all timepoints on YES media revealed strong correlation ($>$0.9) of timepoints $>$20h with the final timepoint (Fig. 1.12C). This is important result means that all timepoints following the rapid growth phase return similar results and there is no 'catching up' of slow growers. This makes precisely defining incubation times unnecessary. Important to note is the (weak) correlation of later timepoints with the initial timepoint, indicating a bias introduced by the amount of inoculation biomass. This has been noted before (Zackrisson et al., 2016) and is due to cell adhesion properties ('stickiness') varying between strains. In practice, this is usually not a problem if the fitness of a strain in a certain condition is compared to the fitness in a control condition, which corrects for this bias. Crucially, later timepoints also correlate extremely



Condition	Correlation: endpoints – max slopes (corrected)
YES	0.959
YES - 0.1% gluc. + 3% glycerol	0.959
YES - 1.5% gluc. + 1.5% glycerol	0.947
YES - 0.5% gluc. + 3% EtOH	0.948
EMM - NH_4Cl	0.976
EMM - Valine	0.957
EMM - Glycine	0.909
EMM - Arginine	0.952

Fig. 1.12: Maximum slopes and endpoints are tightly correlated. **(A)** Hypothetical growth curves, in the form of Gompertz models (Zwietering et al., 1990), illustrating the danger of using arbitrary snap-shots for fitness analysis. Depending on the chosen timepoint, vastly different results can be obtained. **(B)** Real growth curves before (top) and after correction (bottom) of 57 *S. pombe* wild strains (average of approximately 16 biological replicates per strain). Strains were grown on standard, rich YES media and imaged every 20 minutes over a course of 48h. **(C)** Comparison of 144 individual timepoints (grid-corrected and averaged by strain) with each other and with the maximum slope of the growth curve. Individual timepoints are tightly correlated as long as they are taken after the rapid growth phase ends, here at approximately 16h (i and iii). These later timepoints are also tightly correlated with corrected maximum slopes (ii). **(D)** Table of Pearson correlation coefficients of corrected endpoints with corrected maximum slopes across eight different nutrient regimes. Both readouts are consistently and tightly correlated, with a median value of approximately 0.95.
Panels B-D were reproduced from Kamrad et al. (2020b) which is published under a CC-BY license.

well with maximum slopes (r=0.96, Pearson correlation, Fig. 1.11C). Similar values were obtained for the other seven conditions with a median correlation of 0.95 (Fig. 1.12D). This result was somewhat surprising and means that endpoints and maximum slopes essentially contain the same information.

This result is is not trivial and has important theoretical and practical implications. A

growth curve in liquid media is usually used to extract three key parameters: lag phase, maximum growth rate and final biomass. In that case, a single timepoint is influenced by these parameters to varying degrees, depending on when the timepoint is taken. Late timepoints exclusively reflect the nutrient to biomass conversion efficiency, not the growth rate which is usually the main interest. It may seem easy to apply the same concept to growth curves on solid. The key difference between the two however is that colonies grow in close proximity and competition on a common medium. While in liquid cultures each strain has a defined amount of nutrients available to them, colonies on agar keep growing until the medium is (locally) depleted. My interpretation of this data is that instead of each strain having the same amount on nutrients, each strain has roughly the same amount of time available to grow, until the media is exhausted by the ensemble of colonies. This means that colony sizes towards the end of growth specifically reflect growth rates, as the data show.

The focus of a previous study (Zackrisson et al., 2016) using growth curves was on achieving the lowest technical variation possible. This was done by remarkable technological advancements including a calibration of pixel intensities to 'absolute' cell numbers and extractions of population doubling times. *Scan-o-matic* and *pyphe* both achieve very low noise and a power analysis can reveal the number of replicates required to achieve similar statistical power for *pyphe* and *scan-o-matic* (Fig. 1.13B). Indeed, it is in a range where the biological relevance of detectable differences (few percent) can be questioned. Certainly, the biological variance greatly exceeds the technical variance (Fig. 1.13A), which decreases the importance of minimising the latter. More important than the noise is therefore if different readouts contain genuinely unique or more specific information. The increased throughput and lower cost more than offset the higher noise levels in my opinion. We therefore generally tend to use endpoints rather than growth curves and include many replicates.

1.2.3 Colony size and redness are orthogonal and independent readouts

Supported by the *pyphe* pipeline, redness scores are a cost- and time-efficient add-on to traditional colony screens. Phloxine B is simply added to the assay plates and these are scanned once in transmission mode for colony area quantification and once in colour mode for redness analysis. With this '2-for-1' approach, one can obtain two quantitative values from the same colony. But what is the relationship between information contained in colony sizes versus redness scores? To answer this question, 238 *S. pombe* knock-out strains were pheno-

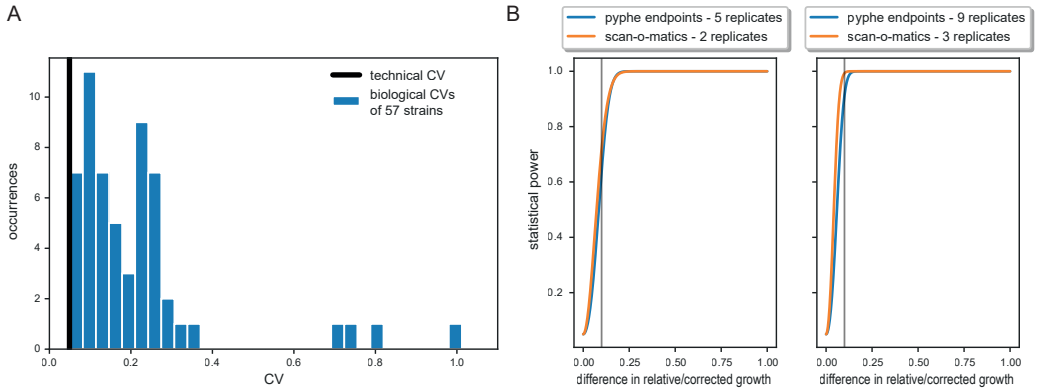


Fig. 1.13: Pyphe is precise enough for the detection of subtle phenotypes. **(A)** Histogram of biological CVs of grid-corrected maximum slopes of 57 wild strains on standard YES media, measured in approximately 16 replicates each. Biological CVs often greatly exceed the technical CV, which was 0.049 (black line). The technical CV was estimated from an additional grid of 96 replicates of the control strain. The key difference between a technical and biological replicate is that technical replicates originated from the same liquid culture which was pinned onto solid agar in 96 format. Biological replicates were kept in separate cryostocks (some also in different multi-well plates), woken up in 384 format and combined with the control strains into 1536 format. **(B)** Comparison of statistical power (chance of false non-rejection of the null hypothesis) between *scan-o-matic* and *pyphe*, assuming a CV of 2% for *scan-o-matic* as published (Zackrisson et al., 2016) and a generous CV of 6% as typically seen for *pyphe* endpoint measurements. More replicates are required with *pyphe* than with *scan-o-matic* in order to achieve a similar power, but this is offset in practice by the increased throughput and lower cost of endpoint measurements.
Panels B was reproduced from Kamrad et al. (2020b) which is published under a CC-BY license.

typed in 70 conditions in triplicates. These genes had been selected (by other lab members) to broadly cover GO-slim categories and also included genes of unknown function. Raw colony sizes and redness scores are weakly anti-correlated (r=-0.19, Fig. 1.14A). This gives some indication that more stressful conditions, which more strongly inhibit growth, also lead to more dead cells within colonies. However this correlation is only weak, suggesting that many conditions inhibit/slow growth without leading to cell death. Technical noise of redness scores within plates followed a strong row-wise pattern (Fig. 1.14B) which was corrected effectively by row/column median normalisation as implemented in *pyphe-analyse*, resulting in a technical CV of only around 1% (Fig. 1.14C). Crucially, after corrections, relative colony sizes and redness score are essentially not correlated (r=-0.088, Fig. 1.14C), indicating that both readouts contain distinct, 'orthogonal' information.

But is there a danger a danger that addition of phloxine B changes the results obtained for colony areas? Colony sizes in seven condition pairs with and without phloxine B were measured and corrected using the *pyphe* pipeline. Unsupervised clustering of the 14 re-

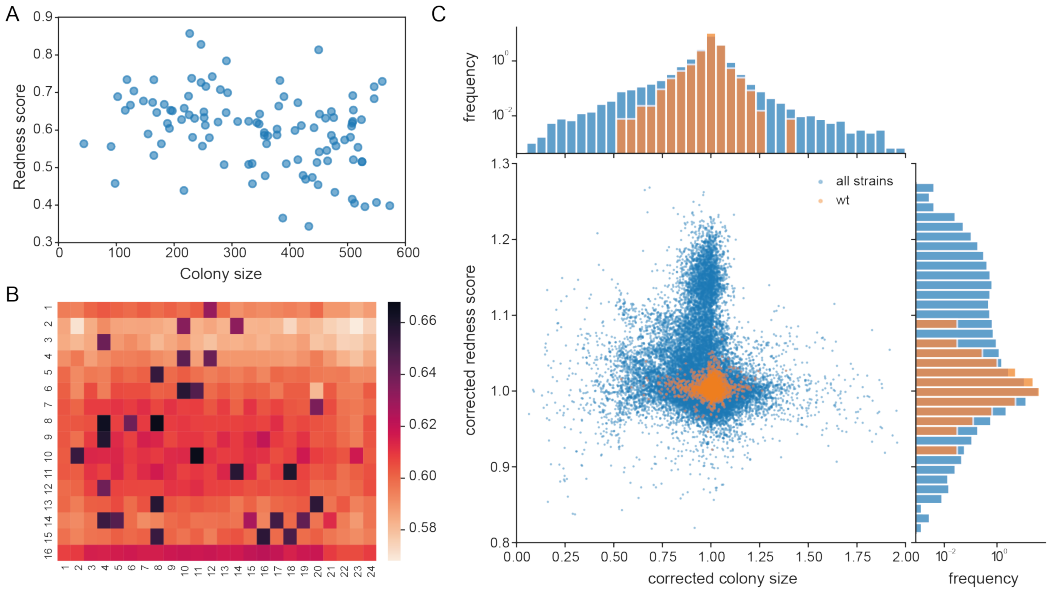


Fig. 1.14: Redness and colony size contain orthogonal information. **(A)** Average redness scores of the standard lab strain 972 in 110 unique batch-condition pairs. Some conditions, such as the standard control conditions, were included in many batches. Each value is the mean of approximately 30 biological replicates. Uncorrected redness scores and uncorrected colony areas are weakly anti-correlated (r=-0.19) **(B)** Median raw redness scores for each physical colony position across 308 plates covering 78 different conditions. There is strong technical variation which follows a row-wise pattern and is an artefact of the imaging method. **(C)** Corrected colony sizes and redness scores of 238 *S. pombe* knock-out mutants relative to the standard lab strain 972 across 110 unique batch-condition pairs. The global CV for redness scores was 1.04% with a FUV of 7.83% (histogram on right). CV and FUV for colony sizes were 6.1% and 31.5% (histogram on top). Both readouts show only very weak anti-correlation (r=-0.088) and therefore provide orthogonal information.
Panels B+C were reproduced from Kamrad et al. (2020b) which is published under a CC-BY license.

sulting conditions showed the condition pairs consistently clustering together (Fig. 1.15A), indicating that the condition is the main driver of the signal and not whether or not phloxine B was included. Condition pairs were measured in different batches which were prepared independently in different weeks. The median correlation of condition pairs was 0.92, a very good value when compared to the median correlation between repeats of the control condition (rich YES media) across different batches (Fig. 1.15B). Differential fitness of individual gene knock-outs with versus without phloxine was assessed by t-tests implemented in *pyphe-interpret*. A single gene, encoding the trehalose-6-phosphate phosphatase Tpp1, showed a significantly decreased colony size ($p_{adj} < 0.05$ and effect size $> \pm\log2(1.05)$) in rich and minimal media (Fig. 1.15C). There is no obvious explanation why particularly this gene should be affected by the addition of phloxine B. More generally, a constant genotype-dependent impact of phloxine B is conceptually not a problem. If phloxine is included in the

assay condition and the control condition to which it is compared, any genotype-specific effects of phloxine B will be corrected for. Conceptually, this is equivalent to adding/changing any other component in the media. This can (and often will) affect growth, which is fine if the media is used throughout an experiment. Overall, phloxine B has no drastic impact on growth of *S. pombe* knock-out mutants and can safely be included in phenotyping experiments.

1.2.4 Conclusions: pyphe enables high-throughput, high-precision phenotyping

This section has demonstrated the power and versatility of the *pyphe* pipeline. It is set up for high-precision, high-throughput colony screens which can detect subtle fitness differences in large libraries of strains in parallel. *Pyphe* has supported numerous colony screens in both the Ralser and Bähler laboratories and will hopefully be a useful tool for the field. It is freely available under a permissive licence from either GitHub or the Python Package index. Besides the tools itself, the comprehensive and complementary test data sets presented have uncovered some general characteristics of colony-based screens which will guide the design of future experiments. First, I have shown that, for arrayed colonies on agar plates, endpoint sizes contain similar information as maximum slopes of growth curves. This result was obtained under conditions designed to break this hypothesis by producing growth dynamics as diverse as possible, namely genetically and phenotypically diverse strains on vastly different nutrient sources. This result is not obvious, assuming a similarity of liquid and solid growth curves, and severely weakens the arguments in favour of using costly, elaborate growth curve experiments. Still, growth curves can provide lower levels of technical noise and *pyphe* cannot quite reach noise levels as low as those published by *scan-o-matic* (Zackrisson et al., 2016). However, compared specifically to *scan-o-matic*, *pyphe* offers a significantly streamlined implementation, eg it does not require a hardware modification of the scanners (incompatible with fire safety regulations), a dedicated local area network (LAN) or image colour calibration strips. The marginally increased level of technical noise obtained with *pyphe* is outweighed by its throughput (which allows more replicates to be measured) and is small compared to the biological noise typically observed in such screens.

Secondly, it has been shown that colony redness scores provide a ‘second dimension’ to colony screens. They accurately reflect the fraction of dead cells in a colony and contain orthogonal, novel information after correction. Given the ease and throughput with

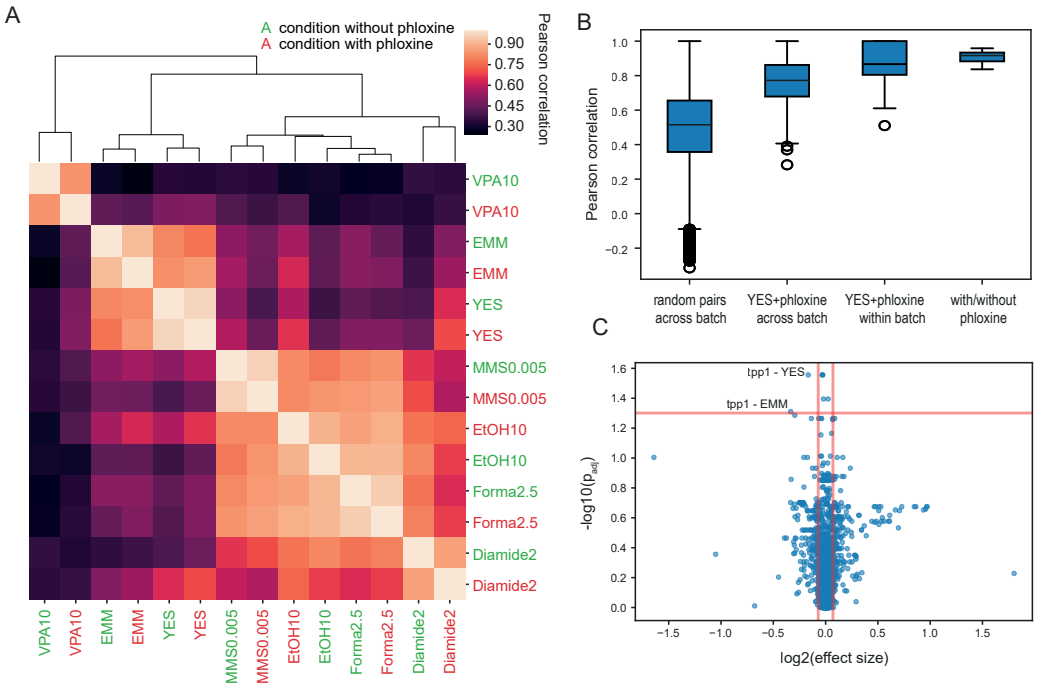


Fig. 1.15: Phloxine B generally does not influence growth. (A) Correlation matrix of seven conditions with and without phloxine B. For each condition, 238 *S. pombe* knock-out mutants were measured in 3 technical replicates. Conditions with and without phloxine B were measured in separate batches in different weeks (with independent plate preparation and revivals of the cryostocks) to avoid batch effects making correlations appear artificially close. Conditions with and without phloxine B consistently cluster together, indicating that the main factor underlying these the obtained scores was the condition, and not whether or not phloxine B was included. Euclidean distance was used as the distance metric and the average method was used for hierarchical clustering. **(B)** Correlation analysis of individual of phenotype scores for 238 *S. pombe* mutants in 110 unique batch-condition pairs, mean of 3 technical replicates each. The control condition YES with phloxine B was included in most batches and the median correlation across batches was 0.77. This is substantially higher than the median correlation of all technical replicate pairs across all conditions (0.51). The median correlation for seven conditions with and without phloxine B (which were also measured in separate batches) was 0.92. It is perhaps surprising that this is even higher than for repeats of the control condition but this is explained of stronger, diverse phenotypic responses triggered by these more stressful conditions. **(C)** Differential fitness analysis of individual genes across the seven conditions with and without phloxine B, performed with *pyphe-interpret*. The effect size is the ratio of the medians of the corrected colony size in media with versus without dye. P-values were computed by Welch's t-test and adjusted with the Benjamini-Hochberg method for controlling the false discovery rate. The chosen effect size cut-off of $\pm\log2(1.05)$ and significance cut-off of $p_{adj} < 0.05$ are illustrated with red lines. A single gene, encoding the trehalose-6-phosphate phosphatase Tpp1, was found to show a significant growth defect in both rich YES and minimal EMM media. These were also the conditions with the most replicates as these were repeated across several batches.

which redness scores can be obtained, this method is a potential game changer, as traditional methods would necessarily require picking and resuspension of colonies followed by flow cytometry, microscopy or CFU counting. Still, much future work is to be done to understand the biological mechanisms behind redness scores and their temporal dynamics. For colony sizes, the data show that the readout remains largely constant if plates are incubated for a few days after rapid growth has ended. This is useful in practice and means incubation times do not need to be precisely defined, certainly there is no 'sweet spot' that one needs to hit. For redness scores the same cannot be expected, as colonies which are left for a long time will show distinct physiological changes including cell differentiations and eventually cell death (Váchová and Palková, 2018). In fact, the observation that relative (corrected) redness scores and colony sizes of *S. pombe* knock-out mutants are uncorrelated is unexpected as colonies with a large fraction fo dead cells should also grow slower. Several explanations for this are plausible, and *pyphe* is ideally set up for future investigations into this matter. One possibility is that cell death is temporally decoupled from growth (ie rapid ageing), another is that cells sacrifice viability for increased growth rate (ie a trade-off where cells effectively go into overdrive, Nakaoka and Wakamoto (2017)) and finally that colonies can tolerate a certain fraction of dead cells as anyway not all cells in a colony are actively dividing (Meunier and Choder, 1999). Resolving these questions may open up new applications for *pyphe* in chronological ageing studies.

Having thoroughly tested and benchmarked the capabilities on test datasets, we set out to analyse the lincRNA knock-out library. Work laid out in this section gave some clear indications of how to do this most effectively and we used endpoints of colony areas (rather than growth curves) and redness scores in a 2D phenotyping approach. The library was screened in 140 conditions, 67 of which also included phloxine B. The experimental part of this project, including image acquisition, was done by María Rodríguez-López and Mimoza Hoti. My analyses, based on the *pyphe* pipeline, revealed 351 condition-specific growth effects in total. Multivariate analysis was performed on the datasets revealing clusters and RNAs and coding genes showing similar responses. Overall, this project revealed numerous new lincRNA-associated phenotypes, providing novel evidence of their function and a rich resource to the community to inform focused studies in the future.

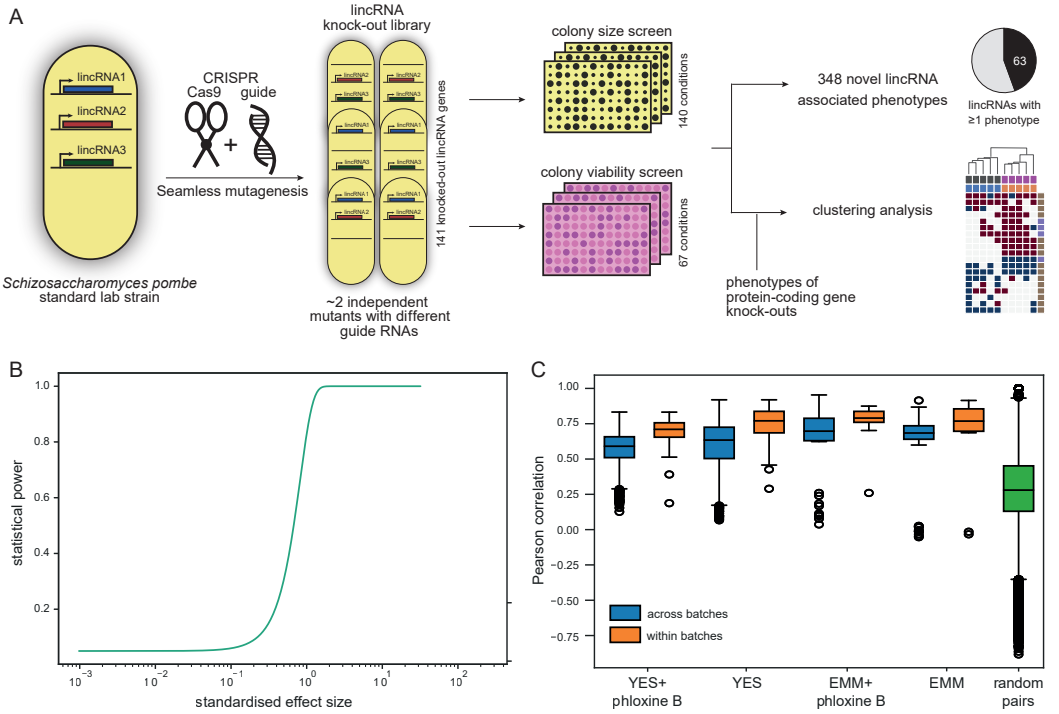


Fig. 1.16: A robust functional genomics workflow to elucidate lincRNA function (A) Overview of functional genomics workflow for characterisation of lincRNA function. A lincRNA knock-out library was constructed using seamless CRISPR/Cas9 genome editing. Multiple guide RNAs were used and several independent successful mutants were kept for each lincRNA gene and the strain library was screened for growth and viability phenotypes in a large number of conditions. The data was analysed to reveal individual, high-confidence hits and global phenotype profile patterns. **(B)** A median of 9 replicates was available for each lincRNA after quality control (and many more for the control conditions), giving our screen the power to detect even subtle phenotypes. Shown is the statistical power (1 - chance of non-rejection of a wrong null-hypothesis) depending on the standardised effect size (difference in means divided by the standard deviation), using a standard variation of 5% as a value typically observed in our data. **(C)** Across 31 experimental batches spread out over several months, repeats of the control conditions correlate only marginally better within batches than across and much worse than random condition pairs, indicating that the dataset is not dominated by batch effects.

1.3 Results

For 141 selected lincRNA genes, multiple independent knock-outs were created in the heterothallic and homothallic versions of the standard lab strain (*972 h-* and *968 h90*). To minimise the risk of off-target effects, almost all lincRNAs were knocked out with more than one guide RNA and usually two or three successful transformants were picked per guide RNA (Fig. 1.16A). The strain library was then arranged in 384 format around a grid of 96 *972 h-* control colonies, compatible with the *pyphe* workflow. Colony growth and viability was assessed in 140 and 67 conditions respectively using three replicate pins per plate. Overall, this resulted in an average of 9 replicates for each lincRNA and each condi-

tion and many more for the control condition. This gave our screen decent power to detect even subtle phenotypes (approximately 5% difference in growth) (Fig. 1.16C).

Standard QC implemented in *pyphe-analyse* was applied to the data, which includes the removal of colonies with circularities above 0.85, data from colonies next to missing grid colonies and the removal of negative and infinite values. Additionally, colony sizes of 0 were set to NA (as these likely originate from pinning errors). Entire plates were discarded if the CV of the internal wild type control was >20% for growth data and >5% for viability data. The experiment was done in 31 batches, spanning several months. This carries a risk of batch effects. For normalisation purposes and to gauge the extent of batch-to-batch variation, two standard conditions (rich YES media and minimal EMM media) were used as control conditions and at least one of each was included in all batches. Although replicates of control conditions within batches correlate more strongly than between batches, the between-batch correlation is still much stronger than the correlation between different conditions, ie the latter dominate the data (Fig. 1.16C).

1.3.1 Novel high-confidence phenotypes for 63 out of 141 lncRNA knock-outs

Using *pyphe-interpret*, I first looked for differences in growth between each lincRNA and the wild type background in the two control conditions (Fig. 1.17A). Hits were called based on a significance threshold of 0.05 after Benjamini-Hochberg correction (across all strains, treating each condition separately) and an effect size threshold of abs(log2(mutant/wild type)) > log2(0.05). This effect size threshold corresponds to the median CV of the dataset and what we consider a biologically noticeable and relevant difference in growth (approximately 5%). In minimal media, 10 lincRNAs were associated with slower growth and one with faster growth. In rich media, 5 lincRNAs were associated with slower growth and none with faster growth. Colony redness analysis in the control conditions identified fewer hits which form a subset of the growth hits (Fig. 1.17B). As effect sizes are generally smaller for redness data, a threshold of abs(log2(mutant/wild type)) > log2(0.015) was used here. Two lincRNAs showed higher redness scores, ie lower viability, on rich media and these had been previously identified as being also slow growers (Fig. 1.17A). A single lincRNA has a lower viability on minimal media. These results confirm our initial working assumptions that lincRNA knock-outs have fewer and weaker phenotypes than coding genes. Indeed many, if not all of these hits, would not have been picked with more traditional methods and would be at most considered a weak phenotype. Furthermore, we could not identify any

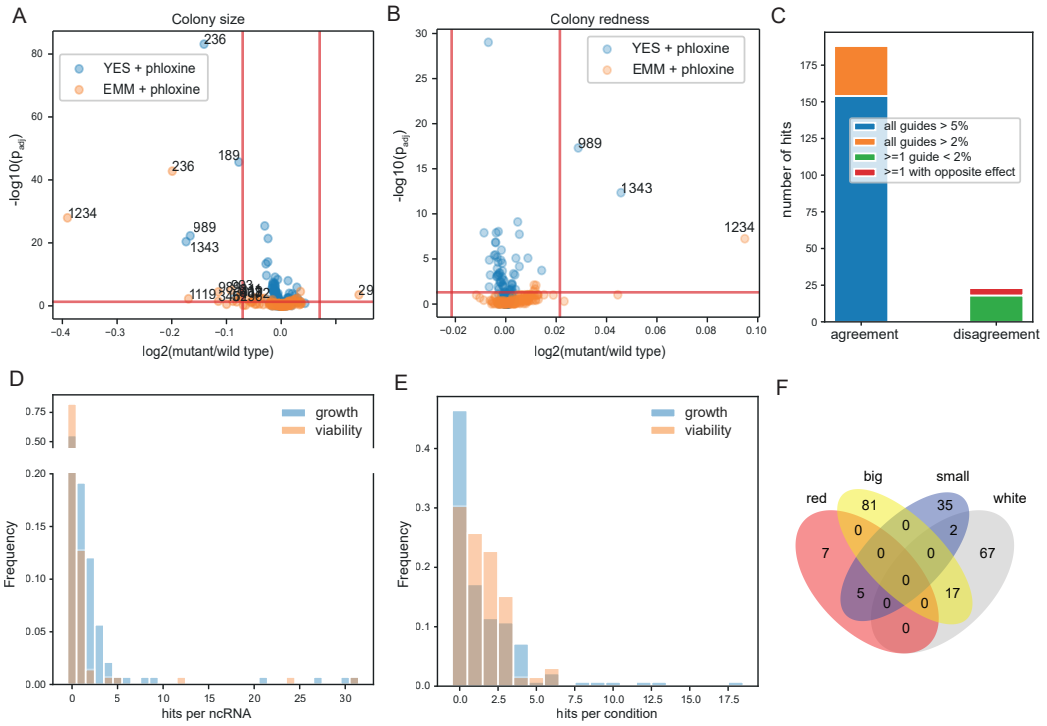


Fig. 1.17: Extensive screening reveals phenotypes for many lincRNAs. **(A)** Volcano plot illustrating differences in growth (approximated by colony area) of lincRNA knock-outs in the two control conditions, rich YES media with phloxine B and minimal EMM media with phloxine B. Thresholds used for hit calling are shown in red ($p_{\text{adj}} < 0.05$ and abs(log2(ratio))>log2(0.05)). **(B)** Volcano plot illustrating differences in viability (approximated by colony redness) of lincRNA knock-outs in the two control conditions, rich YES media with phloxine B and minimal EMM media with phloxine B. Thresholds used for hit calling are shown in red ($p_{\text{adj}} < 0.05$ and abs(log2(ratio))>log2(0.015)). **(C)** Histogram showing the number of hits per lincRNA across 140 conditions (not including the control conditions). Hits were called by comparing the growth and viability of each lincRNA knock-out to that observed in the corresponding control condition, using the same thresholds as in A/B. **(D)** Histogram showing the number of hits per condition across 67 conditions (not including the control conditions). **(E)** Analysis of consistency of phenotypes obtained with different guide RNAs. For the vast majority of lincRNAs considered a hit globally (grouping all data points from different knock-out strains), individual, distinct guide RNAs (usually two or three per lincRNA gene) show good agreement. Guide RNAs are considered to be in agreement if all guide RNAs show a median effect size of $>2\%$ (in the direction of the hit). Only in the minority of cases (24) does one or more of the guide RNAs show no effect ($\leq 2\%$) or an effect in the opposite direction. **(F)** Venn diagram showing overlaps of hits between growth and redness data.
Panel F was prepared by María Rodríguez-López.

essential lincRNAs (which would have resulted in failure to construct the mutant strain by our method).

We next explored condition-specific growth effects using a total of 140 conditions comprising 55 distinct environmental factors (ie many drugs and toxins were applied at different concentrations or media backgrounds). Effect sizes were calculated by taking the ratio of the corrected colony size in the assay condition to the corresponding control condition, thereby focusing on condition-specific effects and correcting for any baseline differences in growth. Using identical thresholds and FDR correction as above, I identified 234 growth phenotypes across all conditions, with 165 resistant and 69 sensitive phenotypes. 63 mutants (44.7%) of lincRNAs has at least one hit (Fig. 1.17D). Most of these had exactly one hit while there were only seven lincRNA genes which significantly altered growth in 5 or more conditions. Just over half (75/145) conditions produced at least one hit (Fig. 1.17E), again with only a small number of conditions producing 5 hits or more.

Hit calling was performed per lincRNA gene (not per individual knock-out mutant strain). I analysed how well data from different guide RNAs agree for those considered a hit globally (Fig. 1.17C). In the vast majority of cases (123 out of 180 hits associated with lincRNA genes represented by at least two independent mutants), all individual guide RNAs agreed and showed an effect size of 5% or more. In 33, all guide RNAs showed at least a similar trend (>2% median effect size). Only in 18 cases one or more of the guide RNAs showed no effect (<2%) and only in 6 cases was there one or more guide with an effect in the opposite direction of the overall hit.

For approximately half of the conditions (67 out of 140), phloxine B was included in the media and plates were also scanned in colour mode. Early batches did not include phloxine B as these were done before this was fully established in the lab. Using the same thresholds as in the control conditions, 25 (17.7%) lincRNAs showed significantly altered viability in at least one condition, relative to the corresponding control condition (Fig. 1.17D). A higher fraction of conditions than in the growth dataset produced at least one hit (46 out of 67) (Fig. 1.17E). Overall, 98 condition-lincRNA pairs showed significantly altered viability. The majority of hits (84) in the redness dataset were resistant phenotypes (higher viability) and 17 out of those were also fast growers (but none slow growers) (Fig. 1.17F). Five condition-lincRNA pairs were also identified as low viability and slow growers. None were fast growers with low viability, compatible with the expectation that colonies with

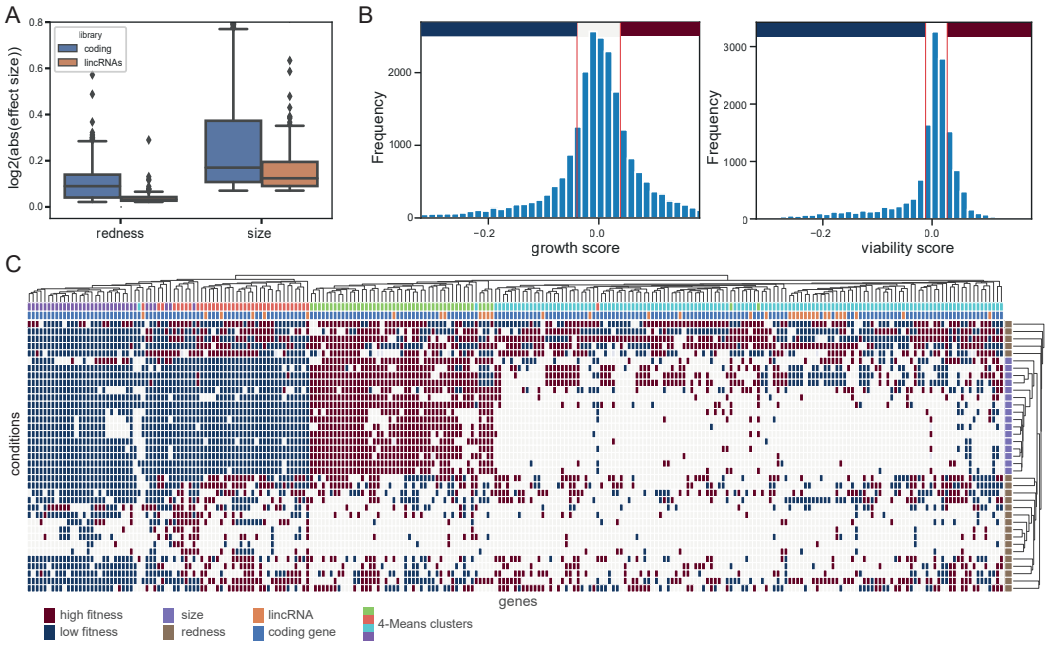


Fig. 1.18: Clustering analysis integrates lincRNA and coding gene phenotypes. **(A)** Absolute effect sizes for redness and size data of (non-essential) coding and lincRNA gene knock-outs. Both libraries were phenotypes in parallel in the same format and conditions, so these data are directly comparable. Coding gene knock-outs show stronger phenotypes compared to lincRNA knock-outs. **(B)** Histograms illustrating median-centred, aggregated growth and viability scores across all conditions and both gene types (lincRNAs and coding). The data was discretised into three groups using the thresholds illustrated in red. Thresholds were chosen symmetrically around 0 so that approximately 33% of datapoints in each dataset were classed as non-zero. **(C)** Clustered heatmap of discretised phenotype profiles of lincRNA and coding gene knock-outs. Shown is data for 37 aggregated core conditions retaining the strongest response across different stressor doses and 249 knock-out lines which showed at least 8 responses across all conditions. Hierarchical clustering was performed using the average method and the Euclidean distance metric was used to quantify profile similarity. Genes were also clustered by 4-means clustering which agreed very well with hierarchical clustering (first row below top dendrogram). All except one k-means cluster contained both coding and non-coding genes (second row from top). The data type (growth or viability) is shown in the first column from the right.

many dead cells should grow slower and only two were slow growers with higher viability. But the majority of hits was only found in either the redness or growth dataset, again highlighting the complementary nature of these two approaches.

1.3.2 Analysis of patterns in knock-out phenotype profiles

In classic functional genomics approaches (eg Mülleder et al. (2016)) genes are grouped into clusters and the function of unknown genes is extrapolated from that of known genes in the same cluster. Since only very little is already known about lincRNA function, clustering only lincRNA profiles will not result in novel functional insights (beyond the observation

that specific sets of lincRNAs might have related functions). We therefore phenotyped 238 non-essential knock-out strains broadly covering different biological processes in parallel with the lincRNA library. I analysed the phenotype profiles of these two groups of strains together, with the goal of associating unknown lincRNAs with known protein-coding gene functions. This multivariate approach relies on the global similarity of profiles and does not require certainty about individual hits, so I only used effect sizes but not statistical significance for this analysis. For many environmental factors, the dataset contained repeats at multiple doses and these were aggregated to a single value, retaining for each gene only the strongest response across all dosages. The resulting value however was still computed as the median across the approximately 9 measured repeats so the risk of accidentally amplifying noise in the data is low. I furthermore filtered out conditions in which only a subset of the mutants was phenotyped (for historical reasons, the strain library expanded while phenotyping was already underway), resulting in 37 core conditions. As expected, coding gene knock-out strains showed on average stronger phenotypes than lincRNA knock-outs (Fig. 1.18A), so I discretised the data to make phenotype vectors of these two gene types more comparable. For both the size and growth data, log2-transformed scores across the entire dataset were median centred and discretised into three groups (low fitness, no response, high fitness) using a threshold that results in approximately a third of the data being non-zero (Fig. 1.18B). These thresholds were $\pm\log2(1.04)$ for size and $\pm\log2(1.02)$ for redness. Genes with fewer than 8 non-zero responses across the 37 conditions were excluded from further analysis, which left 249 mutants, 28 of which lincRNAs. A few (28) remaining missing values in this dataset were 0-imputed and the dataset was clustered using two different methods, hierarchical clustering using the average method and Euclidean distance and k-means clustering with 4 groups. K-means clustering was performed in scipy (Virtanen et al., 2020) using the k-means++ initialisation method and 10 iterations with different centroid seeds out of which the best is chosen. This remediates a key weakness of k-means clustering where the final result depends strongly on the initial centroid position guesses. Both clustering methods agreed very well (Fig. 1.18C), suggesting this is a real signal.

The k-means clusters were used for downstream analysis, because this resulted in groups matching the visual expectation and cutting the dendrogram did not result in suitable clusters. Three of the four clusters contained both lincRNA and coding genes. Cluster 1, contained 32 coding and 6 non-coding genes, including many genes showing increased via-

bility in stationary phase (31.3% of mutants in the cluster), increased resistance to chemical (46.9%), and decreased cell population growth (63.5%). Cluster 2, containing 41 coding and 7 non-coding genes, included long-lived mutants showing increased viability in nitrogen starvation (17.1%) and increased viability in stationary phase (ie glucose starvation, 17.1%). Cluster 3, containing 107 coding and 22 non-coding genes contained a set of lincRNAs situated in the same genomics region. This raises the possibility that the observed phenotypes are driven by cis effects. While this clustering is not fine-grained enough to associate lincRNAs with specific molecular functions or biological processes, it shows that at least a subset of lincRNAs is important for cellular physiology, potentially with regards to stress resistance and longevity.

1.4 Discussion

Pyphe, a novel toolbox for data analysis pipelines for colony fitness screens, has enabled the precise phenotyping of a new library of lincRNA knock-out strains. This was achieved mainly through the implementation of current state-of-the-art normalisation procedures and the use of complimentary colony viability phenotypes. The experimental design, together with the low technical noise of our phenotyping pipeline allowed the detection of differences in growth as low as 5% with reasonable statistical power. Overall, we detect 351 unique hits, covering approximately half of the lincRNAs in the library. I believe this represents a significant achievement, showing that these genes do generally indeed serve biological roles. However, given the lack of (condition-specific) essentiality or even very strong growth defects, these data indicate that lincRNA fine tune cellular phenotypes and that cells on average depend on them less than on protein-coding genes. The performed clustering analysis provided interesting insights and showed that a subset of lincRNA genes shows similar phenotype patterns to coding genes. These lincRNAs, with the most and the strongest phenotypes should naturally be prioritised for further studies. Unfortunately, the clustering analysis did not provide clusters specific enough to annotate lincRNAs to particular cellular functions or GO terms. One reason for that is that only a small subset of non-essential coding genes was included in the dataset, many of them with unknown function. If phenotype data for all coding genes were available, the specificity and power of clustering would improve. The other problem lies in the nature of the data itself. Phenotype profiles are overall very sparse and effect sizes small. This is consistent with the notion that cellular growth is buffered and insensitive to many molecular perturbations (Poyatos,

2020). The use of molecular phenotypes, such as gene expression or amino acid profiles might therefore be a more suitable approach to generate data for these types of clustering analyses (Uygun et al., 2016; Mülleder et al., 2016).

Nonetheless, I believe this dataset will be a valuable resource to the *S. pombe* community, serving as a starting point for other large-scale projects using the novel lincRNA library as well as for more focused projects. It is well suited to derive hypotheses about the function of specific lincRNAs or which lincRNAs might be important for a specific function (related to one of the conditions in the dataset). A key challenge for future studies will be to decipher the molecular mechanism by which these lincRNAs exert their effect on the phenotype. Cis-effects, where lincRNAs change the transcription of neighbouring genes, are a likely contributor but the relative importance of cis and trans effects remains unclear (Balarezo-Cisneros et al., 2020).

Chapter 2

A natural pyruvate kinase variant affects growth and stress resistance in *S. pombe*

A well-established paradigm is that Crabtree-positive yeasts like *S. pombe* suppresses respiration in the presence of glucose and convert sugar to ethanol by fermentation. However, recent studies have highlighted the importance of respiration, even during growth on glucose, and paint a more nuanced picture where cells balance the activity of both pathways according to their metabolic needs. I set out to investigate the genetic basis of this respiration-fermentation balance, using a panel of *S. pombe* wild isolates. Treating it as a complex, quantitative phenotype, genome-wide association linked the residual dependence on respiration to a single nucleotide polymorphism in *pyk1*, *S. pombe's* only gene encoding a pyruvate kinase. The mutation was found to alter glycolytic flux, with the high-activity isoform leading to more fermentation and faster growth. This was associated with wide-ranging changes in the transcriptome and proteome, as well as other phenotypes such as chronological lifespan and altered resistance to a range of chemical compounds. Most notably, the high-activity isoform drastically lowers the resistance to oxidative stress and might therefore represent an ecological trade-off between growth and stress resistance. Importantly, the *S. pombe* standard lab strain was found to carry the low-activity isoform, which is the minor allele among wild strains and not found in other known eukaryotes. This metabolic particularity of the reference strain, and the associated phenotypes, could have important implications for future research using this key model organism.

2.1 Introduction

Glycolysis is an ancient metabolic pathway (Ralser, 2018) and the backbone of carbon metabolism in most organisms and cell types. While the involved metabolic intermediates,

the sequencing of reactions and the catalysing enzymes are the topic of much of the classical biochemical literature (Barnett, 2003), the regulation of glycolysis and its integration into the systems context remain active areas of investigation (eg Hackett et al. (2016); Zelezniak et al. (2018)). Glycolysis provides energy in the form two net ADP $\rightarrow$ ATP conversions and glycolytic intermediates feed into several other key pathways. One of these, the pentose phosphate pathway (PPP), generates pentoses for nucleotide biosynthesis and reducing equivalents in the form of NADPH. Simultaneously, many amino acids are produced from glycolytic intermediates (Ljungdahl and Daignan-Fornier, 2012). The end-product of glycolysis is pyruvate which in yeast is further metabolised by two pathways, the mitochondrial TCA cycle which provides reducing equivalents for energy production by oxidative phosphorylation ('respiration') or alternatively fermentation to form ethanol. Ethanol is secreted from cells as a waste product, referred to as overflow metabolism (Paczia et al., 2012). This makes glucose utilisation by fermentation appear inefficient, as more than ten times as many ATP molecules can be generated by full oxidation of glucose through the TCA cycle and the electron transport chain (Rich, 2003).

Respiration and fermentation therefore represent two distinct modes of carbon metabolism and the two are often antagonistically regulated (Molenaar et al., 2009; Takeda et al., 2015). Some cells prefer fermentation over respiration in glucose-replete conditions, even if oxygen is available. This is (somewhat confusingly) called aerobic glycolysis (Crabtree, 1929) and is observed in microbial species which are classified as Crabtree-positive. Analogously, mammalian cancer cells generally exhibit aerobic glycolysis, which is known as the Warburg effect (Warburg, 1927). In microbes, aerobic glycolysis is generally associated with fast growth and it is thought to have been selected for based on its ability to provide high ATP production rates (Pfeiffer and Morley, 2014). However, the mechanistic explanation for how aerobic glycolysis can support faster growth compared to respiration remains an active area of investigation. Most likely, the key lies in the observation that the molecular machinery required to catalyse both pathways are vastly different and under heavy cellular constraints. Respiration requires many more individual proteins and protein complexes. These in themselves are costly to produce, making energy production by fermentation the overall more efficient allocation of resources (Basan et al., 2015; Mori et al., 2019). Alternatively, or simultaneously, respiration requires large complexes integrated into the mitochondrial membranes. Membrane space restrictions will therefore represent an

upper limit to cellular respiration flux (Andersen and von Meyenburg, 1980; Zhuang et al., 2011; Vazquez and Oltvai, 2016; Szenk et al., 2017). However, *S. cerevisiae* and *S. pombe* cells actively repress respiration genes (and many other genes) in glucose-rich conditions, known as glucose repression or carbon catabolite repression (Ronne, 1995; Vassiliadis et al., 2019), which is somewhat inconsistent with this 'membrane real-estate hypothesis'.

The classical dogma in which Crabtree-positive species, such as the model yeasts *S. pombe* and *S. cerevisiae*, solely rely on fermentation is recently replaced by a more nuanced understanding which considers the simultaneous roles of central carbon metabolism in energy (ATP) generation and provision of biosynthetic intermediates. A key observation in this direction is that Crabtree-positive yeast species still require some respiration activity for optimal growth on glucose. *S. cerevisiae* can grow without mitochondrial genome (and thus respiration) but this results in a growth defect, known as *petite* phenotype (Ephrussi, B. and Hottinguer, H. and Tavlitzki, J., 1949). *S. pombe* cannot normally grow without its mitochondrial genome (Haffter and Fox, 1992; Heslot et al., 1970; Chiron et al., 2007). But also in this species, blocking the electron transport chain (with the antimycin A) has been shown to lead to a moderate growth defect on rich media and a strong growth defect on minimal media with ammonium as the sole nitrogen source (Malecki et al., 2016). We recently traced down this growth defect to cells becoming unable to produce amino acids derived from alpha-ketoglutarate (Malecki et al., 2020) and a similar link has been observed in *S. cerevisiae* (Druseikis et al., 2019). A more nuanced picture therefore emerges in which cells finely balance respiration and fermentation in consideration of efficient energy production and required metabolic building blocks.

I set out to investigate the genetic basis of this respiration-fermentation balance by treating it as a complex, quantitative phenotype, approximated as the amount of residual respiration under glucose-rich conditions. Based on our collection of wild *S. pombe* isolates (Jeffares et al., 2015), a genome-wide association study (GWAS) identified a miss-sense variant in the *pyk1* gene, encoding *S. pombe's* sole pyruvate kinase (PYK). PYK catalyses the conversion of phosphoenolpyruvate (PEP) to pyruvate, the final step of glycolysis. It has been previously described that PYK flux is linked to respiration/fermentation balance (Yu et al., 2018; Grüning et al., 2011), but the ecological relevance and downstream impact of this relationship has not been described. Using CRISPR/Cas9-engineered allele replacement strains, the impact of this variant on metabolome, transcriptome, proteome, physiol-

ogy and fitness is dissected in detail. The allele found in the minority of strains, including the standard laboratory strains used by all researchers, is found to lower PYK activity and thereby glycolytic flux. This leads to a wide-ranging reconfiguration of gene expression and results in slower growth but increased resistance to several stresses. This work uncovers a metabolic peculiarity of the sole set of reference strains used throughout the *S. pombe* community. Furthermore, it suggests an adaptive role of the *pyk1* mutation in regulating stress resistance and highlights the general importance of central carbon metabolism for intially seemingly unrelated phenotypes such as stress resistance.

2.2 Results

2.2.1 GWAS implicates *pyk1* in regulation of 'Crabtreeness'

As outlined above, the Crabtree effect is usually regarded as being of qualitative nature, with species either classified as Crabtree-positive or -negative. *S. pombe* is classed as Crabtree-positive but depends on respiration to some degree, which challenges this binary classification principle. I wondered whether there could be intra-specific differences in the 'Crabtreeness' within wild *S. pombe* isolates and if the Crabtree effect could thereby instead be viewed as a quantitative, complex phenotype. This can be quantified practically by measuring the extent to which strains still respire even though they are growing on glucose-rich media (residual respiration). Respiration rates of yeast cultures have been measured in various ways in the laboratory, usually in the form of oxygen consumption rates. Our collection of wild strains (Jeffares et al., 2015) comprises over 100 strains which required a technique that can be applied in reasonably high throughput. I therefore decided to use a proxy readout: A strain grown on glucose media containing an inhibitor of the respiratory chain will grow slower than the same strain on media without the inhibitor, and crucially, the magnitude of that growth difference reflects the dependence on residual respiration. Antimycin A is a suitable inhibitor which blocks ubiquinol-cytochrome c oxidoreductase (Kim et al., 1999). We have previously used antimycin in this species and showed that the standard laboratory strain *972 h-* has a moderately reduced growth rate and biomass yield upon treatment (Malecki and Bähler, 2016; Malecki et al., 2016). This proxy readout can be obtained from a colony size assay, which I implemented using the *pyphe* toolbox (Kamrad et al., 2020b).

The fitness of 154 strains was determined on rich glucose (YES) media with and without antiymycin (Fig. 2.1A). The wild strain library was re-arranged to conform to the recom-

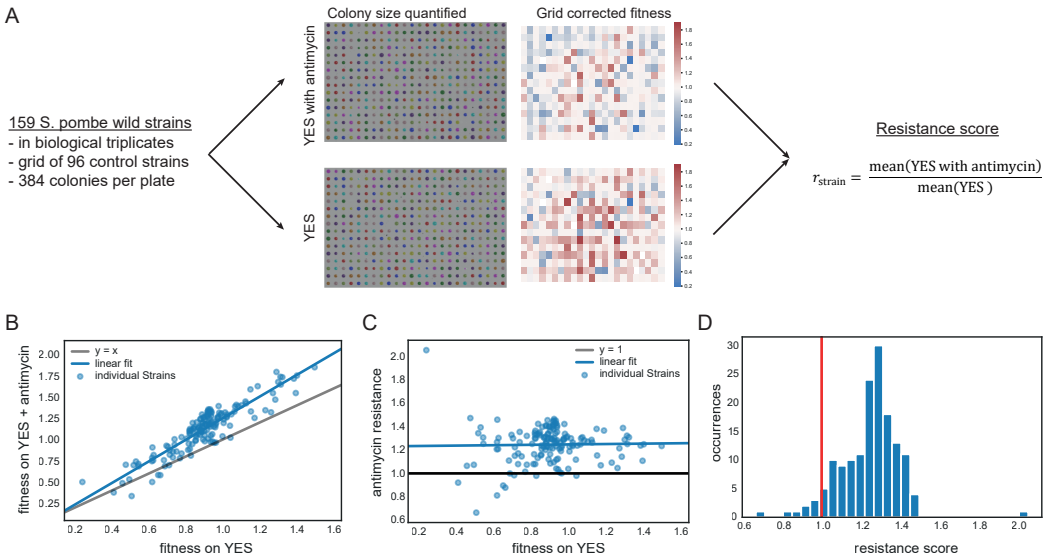


Fig. 2.1: Quantification of antimycin resistance of *S. pombe* wild strains. **(A)** Schematics of phenotyping experiment. 2 plates were prepared by randomly arranging the 159 wild strains in triplicates around a grid 96 colonies of the control strain *972*. These plates were copied onto media with and without 500µg/L antimycin A and scanned after rapid growth had ended. Colony sizes were quantified and corrected for spatial effects using the reference grid. After quality control (see Materials and Methods), resistance scores for each strain were calculated by averaging corrected colony sizes in each condition and taking the ratio of media with versus without antimycin A. **(B)** Scatter plot showing average corrected colony sizes for each strain on media with versus without antimycin A. Both variables are strongly correlated (R^2 = 0.83), indicating that the strongest factor explaining the observed fitness on media with antimycin is not antimycin resistance but rather basal growth levels in the given setting. The equation of the fitted line is y = 1.27x - 0.02. Due to the nature of the normalisation procedure, the standard lab strain has a fitness of 1 in both conditions. The slope of the fitted line is not equal to 1 (black y=x line) due to the control strain falling into different areas of the distribution in the two conditions. **(C)** The basal growth difference can be accounted for by taking the ratio of both variables, which yields the antimycin resistance score. After forming the ratio, strains show a range of resistance scores which are not correlated with basal growth. **(D)** Histogram of resistance scores which are relative to the resistance of the standard lab strain *972* marked with a red line. The mean resistance across all strains is 1.25±0.15.

Panel B and D are reproduced with minor changes from Kamrad et al. (2020a) which is published under a CC-BY license.

mended *pyphe* format with 384 colonies per plate. Individual strains were arranged in triplicates around a grid of 96 control strain (*972*) colonies with additional interspersed controls. Colony sizes on media with and without antimycin were corrected for within- and between-plate variation using the reference grid correction and averaged for each strain. Additional quality control filters were applied to filter out spurious data (see Materials and Methods). Fitness on media with and without antimycin was correlated (R^2 = 0.83, Fig. 2.1B). This indicates that a large (in this case uninteresting) component of the colony size is explained by a general growth effect, reflecting how well strains perform in the given (laboratory) setting. Antimycin resistance was determined as the ratio of the mean corrected colony size in medium with versus without antimycin, which is not correlated with basal fitness on YES (Fig. 2.1C). By the nature of the normalisation procedure, where fitness is reported relative to the control strain, the standard lab strain *972* had a resistance score of 1. But perhaps somewhat surprisingly, it ranked among the most susceptible strains (rank 10 of 154). The mean resistance score across all strains was 1.25, but with a substantial range of values (standard deviation = 0.15) (Fig. 2.1D).

Next, I investigated if the observed range of antimycin resistance scores across different wild isolates could be due to the genetic differences between the strains. The collection of strains contains considerable genetic diversity, with some lineages having diverged approximately 2000 years ago and a total of approximately 180,000 called genomic variants (Jeffares et al., 2015). Narrow-sense heritability (h) is a popular measure that can be computed from a matrix of genomic variants and the phenotype values and that reflects the proportion of the phenotypic variance explained by additive genetic effects (Zuk et al., 2012). The estimated h for antimycin resistance was 0.54. This is substantially higher than for most of the 223 phenotypes previously reported, where the median was 0.29 (Jeffares et al., 2015). I next identified individual variants statistically linked to the phenotype using a genome-wide association study (GWAS) (Fig. 2.2A). Variants were called from published whole-genome shotgun sequencing data (Jeffares et al., 2015) using *freebayes* and only high-quality, suitable variants (118,527 in total) were used for GWAS (Materials and Methods).

GWAS was performed by mixed-model regression, using a linkage-disequilibrium adjusted kinship matrix to correct for population structure. P-values for each variant are shown in Fig. 2.2B. The small size of our strain collection limited the statistical power of associ-

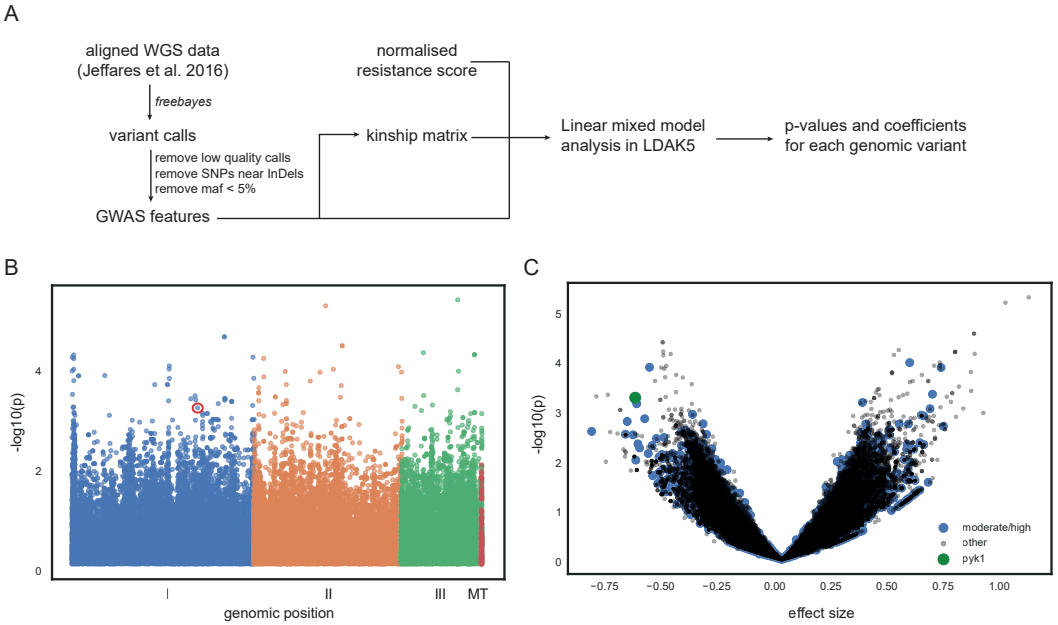


Fig. 2.2: GWAS identifies potentially causal variants. **(A)** Small genomic variants were called using *freebayes* and filtered for high-quality variants with a minor allele frequency above 5%. These were used to construct a kinship matrix in *LDAK5* and a linear mixed model regression was performed on normalised phenotypes (transformed to normal shape, centred around 0 and scaled to variance 1) using *LDAK5*. **(B)** Statistical significance of each variant plotted against genomic position. **(C)** Additionally to the effect size, the p-value depends on other factors, such as the minor allele frequency. Therefore, besides statistical significance (y-axis), the size of the model coefficient (x-axis) was also considered during interpretation of GWAS results. To further narrow down on promising hits, only variants with predicted moderate or severe impact (obtained with *SnpEFF)* were considered. One of the highest scoring variants is a miss-sense mutation in *pyk1*.
Panel C is reproduced with minor changes from Kamrad et al. (2020a) which is published under a CC-BY license.

ations. The predicted variant impact, obtained with *SnpEff* (Cingolani et al., 2012), was therefore used to to prioritise hits for follow-ups based on other criteria (Fig. 2.2C). Among the top-100 most strongly associated variants, were only 8 with moderate or high predicted impact. 2 of those were located in genes with no functional annotation in PomBase (Wood et al., 2012) at the time, and another 2 were located in *S. pombe* specific genes, *wtf16* and *wtf8* (Hu et al., 2017). The remaining 4 variants were located in in *pfl5*, encoding a cell-surface glycoprotein, *jac1*, encoding a mitochondrial 2Fe-2S cluster assembly co-chaperone and *ubp9*, encoding a ubiquitin C-terminal hydrolase. Particularly striking was an associated SNP (p=0.00076) in a gene encoding the key glycolytic enzyme Pyk1. The T→C variant at the genomic locus I:3845516 that leads to a T→A amino acid substitution at residue 343 of the protein sequence.

Strains with the non-reference allele (‘A-allele’) showed a 15% higher median resis-

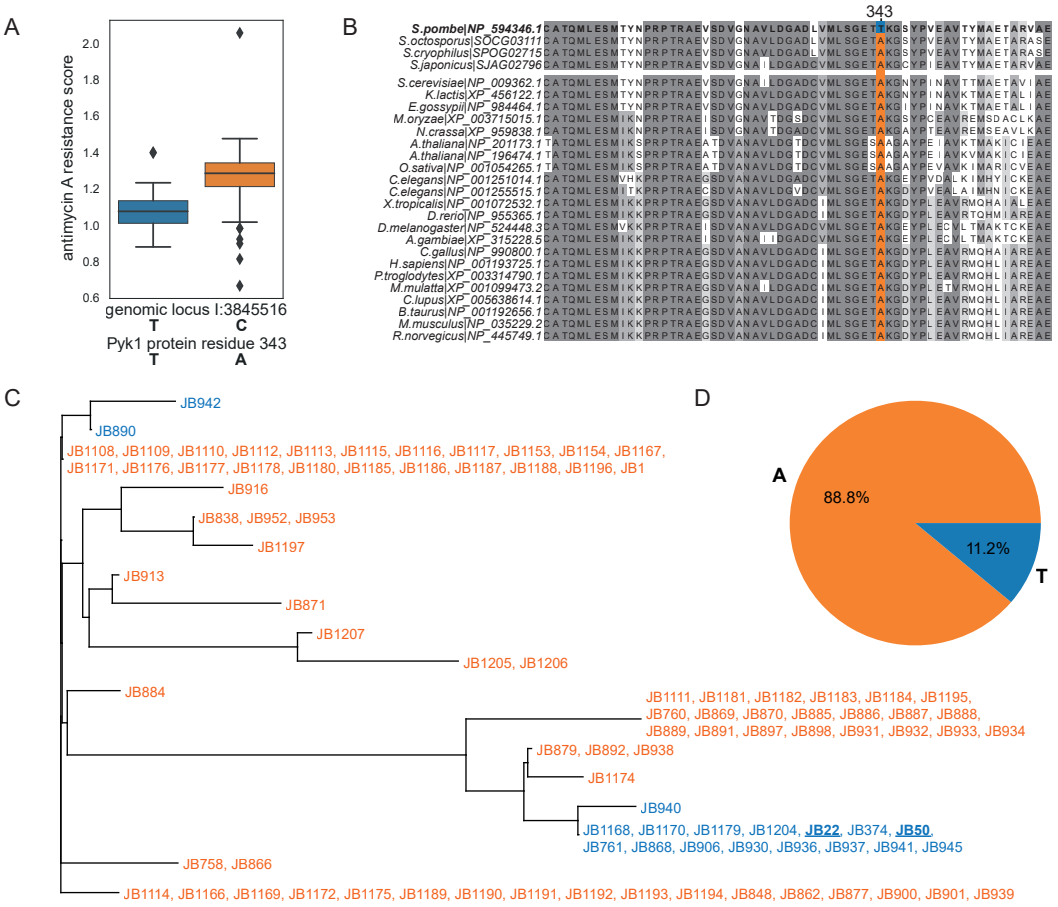


Fig. 2.3: An unusual and rare miss-sense variant in *pyk1*. **(A)** Antimycin resistance of 154 strains grouped by *pyk1* allele. Strains with a cytosine at this locus (which results in an alanine residue in the protein sequence) tend to have a higher resistance score than strains with a thymine (which results in a threonine residue in the protein). **(B)** Protein sequence alignment of the region surrounding the mutation of interest. Darkness of shading reflects degree of conservation at the position. The *S. pombe* reference sequence is the only protein with a T in position 343, out of 25 homologues from distant eukaryotes and 3 close homologues from the *Schizosaccharomyces* genus. **(C)** Phylogenetic analysis of the *pyk1* gene and immediately neighbouring regions with strains carrying the T- and A-allele in blue and orange respectively. Strain identifiers are the same as in Jeffares et al. (2015). The heterothallic (JB22) and homothallic (JB50) varieties of the standard lab strain are underlined and bold. **(D)** Allele frequency at the locus across all wild strains.

tance compared to strains carrying the reference allele ('T-allele') (median resistance scores of 1.28 versus 1.07, Fig. 2.3A). The *S. pombe* reference genome, unlike other eukaryotes, only contains a single gene encoding a PYK enzyme. I also found no evidence for additional PYK enzymes while searching published *de novo* assemblies (Jeffares et al., 2015) for Pyk1 homologues using *tblastn* (Camacho et al., 2009). We (Jan Grossbach and I) investigated the evolutionary conservation of the region surrounding the mutation of interest and collected homologous protein sequences from three other members of the same genus and a range of other eukaryotes. Aligned sequences were strongly conserved (45-93% agreement with the consensus sequence called from the alignment). The 6 residues preceding and 2 residues following the position of interest were especially strong conserved, with only *Arabidopsis thaliana* deviating from the consensus (Fig. 2.3B). This region is described as part of the binding pocket for one of the substrates, ADP (Schormann et al., 2019). Crucially, the *S. pombe* Pyk1 sequence, which is based on the reference genome carrying the T-allele, is the only one with a threonine in position 343. Together, these observations led us to hypothesise that the mutation affects the protein function which result in a lower 'Crabtreeness' of the standard lab strain and other strains carrying the unusual T-allele.

Next, I investigated the distribution of A- and T-alleles within the *S. pombe* wild strain collection. The reference T-allele was the minor allele found in 18 of 161 strains (Fig. 2.3D). I constructed a phylogenetic tree from bi-allelic SNPs in the genomic region containing the *pyk1* gene ($\pm$ 500bp up- and down-stream) using the neighbour joining method (Fig. 2.3C). For several sets of strains, members were indistinguishable based on bi-allelic SNPs from that region and therefore grouped in the same leaf. Overall, four unique sequences carried the T-allele and these were found in two highly distant lineages. The T-allele is also found in distantly related lineages considering a whole-genome consensus tree, which can be found in Jeffares et al. (2015). A simple and reasonable explanation is that the T-allele arose twice independently in these lineages from an ancestral A-allele.

The strain collection contains some meta-data describing when and where strains were isolated (Jeffares et al., 2015). I wondered whether the ecological niches from which strains carrying the T-allele ('T-strains') were isolated could give any indication regarding its function. Out of the 18 strains, six were isolated from Europe, one from Asia and one from Australia. For the remaining 10, the geographical origin is unknown. *S. pombe* is usually isolated from human-created but non-industrialised sugar-rich, spontaneous fermenta-

tions (Jeffares, 2018). Accordingly, the predominate substrate of T-strains were fermenting grapes. A single strain was isolated from lychee fruits and another from glace syrup. Neither the geographical origin nor the niches from which they were isolated suggest clear hypotheses for why these strains should have different respiration dependence or glycolytic metabolism. However, these niches and locations do not necessarily reflect past selection regimes, as the global dispersal of *S. pombe* is relatively recent (Jeffares et al., 2015) and the natural origin of strains found in these fermentations is unclear.

2.2.2 Metabolomic analysis of allele swap strain reveals lower PYK activity in T-strain

Next, we (María Rodrígues-López, Jan Grossbach, Michael Muelleder, Valentina Capalletti and I) characterised the impact of the mutation on metabolism and gene expression. Wild strains are of limited use for studying the impact of this mutation. Every strain contains thousands of mutations and displays different growth behaviours (eg flocculation) and morphology, making it impossible to isolate the effect of the *pyk1* allele from effects of and interactions with the genetic background. Using CRISPR/Cas9 gene editing (Rodríguez-López et al., 2016), allele replacement strains were constructed in the standard laboratory strain *h-* and *h90* backgrounds. The standard laboratory strains *972 h-* and *968 h90* are from here also referred to as T-strain and the allele replacement strains with the $Pyk1^{T343A}$ mutation as A-strains. Three successful transformants were kept and used as biological replicates. This method allowed us to manipulate a single base pair 'seamlessly' in the genome without introducing additional genetic perturbations, such as resistance cassettes, which could confound results.

A targeted metabolomics workflow based on liquid chromatography and tandem mass spectrometry (LC-MS/MS) was used to quantify key central carbon metabolism intermediates covering glycolysis, pentose phosphate pathway (PPP), tricarboxylic acid (TCA) cycle, redox co-factors and adenine nucleotides. Metabolites were quantified using external calibration for 9 replicates of the T-strain and 8 replicates of the A strains. I corrected metabolite concentrations in each sample for the OD of the culture at the time of sampling. 9 cultures were sampled twice and processed as technical duplicates. Across all metabolites and technical replicate pairs, OD-corrected concentrations correlated extremely well (r=0.93, Pearson correlation, Fig. 2.4A). The median CVs across all metabolites were 0.15 and 0.18 in the T- and A-strain respectively (Fig. 2.4B). indicating that the sample preparation and

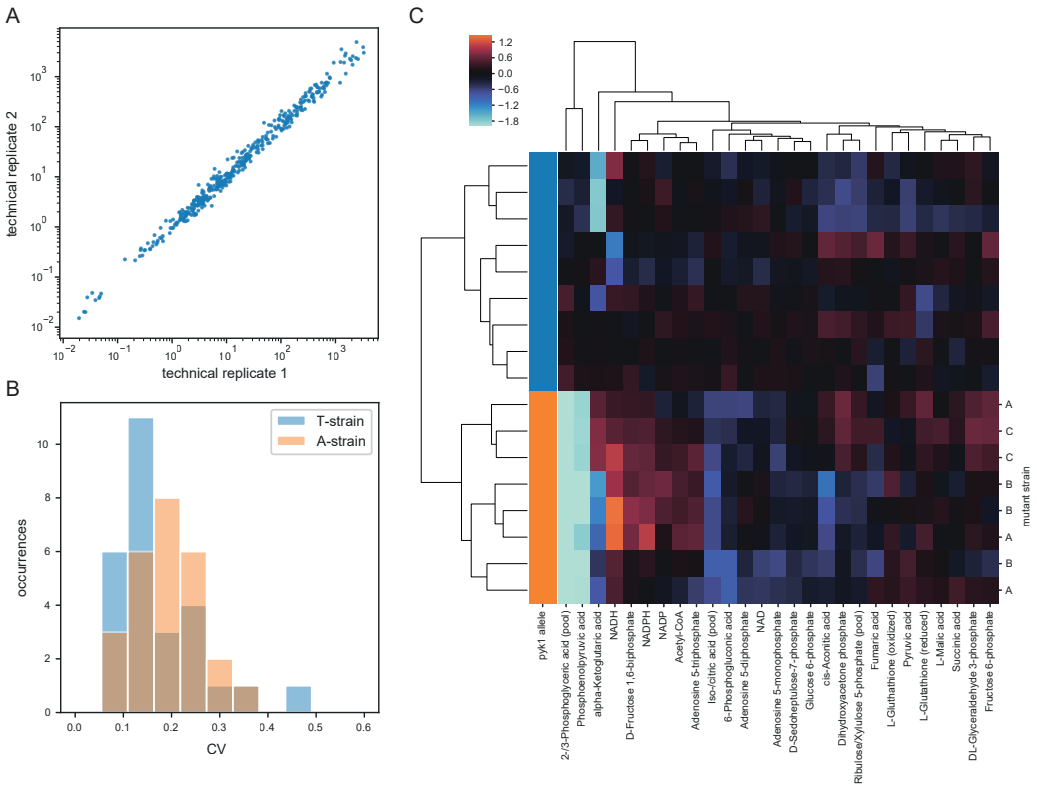


Fig. 2.4: Metabolomic profiling of allele swap strain by LC-MS/MS. **(A)** From 9 cultures, 2 aliquots were harvested and prepared in parallel as technical replicates. Shown are all individual OD-corrected metabolite measurement pairs across the 9 cultures. The overall correlation is 0.93, indicating that our sample preparation and quantification workflow is robust and reproducible. **(B)** Distribution of CVs obtained from biological replicates for individual metabolites. The median CV was slightly higher for the A-strain compared to the T-strain (0.18 versus 0.15). **(C)** Heatmap and hierarchical clustering of metabolome data of T-and A-strains. For each metabolite, OD-corrected concentrations were divided by the median of the T-strain and log2 transformed. This corrects for general abundance differences between metabolites but preserves the relative variation for each metabolite. To aid visualisation, the colour bar was clipped at -2. Euclidean distance and the average method were used for clustering which clearly distinguishes strains based on their *pyk1* allele (shown in first row, orange and blue denote A- and T-strain respectively).

quantification methodology performed well. Unsupervised clustering of the dataset clearly distinguished strains based on their *pyk1* allele (Fig. 2.4B), indicating that the allele-swap has a marked impact on the overall metabolic profile. Importantly, the three independent A-strain transformants (denoted A,B and C) did not cluster together, which would have been indicative of inconsistencies during mutagenesis (such as off-target effects).

The two most striking difference in the metabolic profile of the A-strain compared to the T-strain is the strong depletion of glycolytic intermediates upstream of PYK (Fig. 2.4C). Phosphoenolpyruvate levels were only 25.9% of the T-strain level, for 2- and 3-phosphoglyceric acid (which were indistinguishable by our analytical method) the same value was even lower at 12.7% (Fig. 2.5A, p_{adj}=8.7x10^{-8} and 3.7x10^{-6}, Welch's t-test, corrected for multiple testing across the entire metabolomics dataset with the Benjamini-Hochberg method, t-tests are unpaired and two-sided throughout this chapter). This depletion could be explained by an increased activity of Pyk1, which would be consistent with the earlier observations regarding the location of the mutation in the protein structure. It has long been known that *S. pombe* has lower PYK activity than *S. cerevisiae* (Nairn et al., 1995, 1998), but this of course refers to the standard lab strain which carries the T-allele. I tested directly whether the A-allele enabled higher PYK activity using an enzyme activity assay. In this assay, previously used by others (Gehrig et al., 2017), the PYK reaction (for which the substrates are supplied in the buffer) is coupled to the reduction of pyruvate to lactate (by purified lactate dehydrogenase, which is supplied in excess). For every molecule of pyruvate produced, a molecule of NADH (also contained in the buffer) is consumed and this decrease in concentration can be followed colourimetrically by absorption measurements at 340nm (Fig. 2.5B). NADH concentration timecourses were recorded for 3 biological replicates of both strains, each measured in technical duplicate assays (Fig. 2.5C). After determination of slopes of each trace, blank subtraction and averaging of the technical repeats, a 87% higher mean activity was obtained for the A-strain (Fig. 2.5D, p=0.0072, Welch's t-test). The quantitative absolute values should not be over-interpreted as activity will depend in the details of the buffer composition and activity in the assay does not necessarily equal activity *in vivo*. For example, the enzyme might become activated or inactivated during lysate preparation. To specifically prevent the inactivation of Pyk1 by oxidation, DTT was included in the lysis buffer. Furthermore, this single result is not a thorough enzymological characterisation of the A- and T-proteins, which would have required purified proteins. But together with the

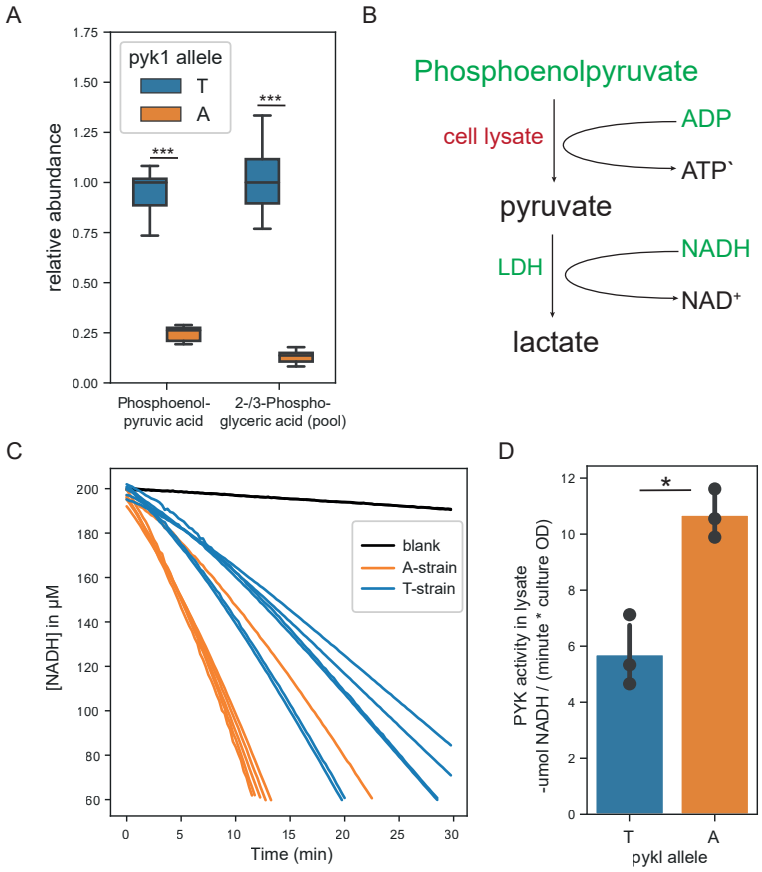


Fig. 2.5: The A-strain features higher PYK activity. (A) Metabolomic analysis shows a strong depletion of glycolytic intermediates just upstream of Pyk1. Phosphoenolpyruvate levels and 2-/3-phosphoglyceric acid levels were 25.9% and 12.7% of the levels of the A-strain. (p_{adj}=8.7x10^{-8} and 3.7x10^{-6}, Welch's t-test). **(B)** PYK activity in strains was measured using a lactate dehydrogenase (LDH) coupled photometric assay. Cell lysate is added to a pH-buffered reaction mix which contained metabolites and purified LDH (green) as well as potassium and magnesium ions. The produced pyruvate is then reduced to lactate by LDH which is present in excess which results in equimolar consumption of NADH (the concentration of which is observable through absorbance measurements at 340nm). **(C)** NADH concentration traces obtained for 3 biological replicates of both strains, each measured in technical duplicate assays. **(D)** Lines were fitted to each trace, blank-subtracted and the two technical repeats were averaged. Activities correspond to the average slope divided by the OD of the culture at the time of sampling. Shown are mean OD-corrected PYK activities in lysate of both strains (± standard deviation), based on data shown in (C).
Significance keys: *p $<$ 0.05, **p $<$ 0.005, ***p $<$ 0.0005 (Welch's t-test). Panel A and D were reproduced with minor changes from Kamrad et al. (2020a) which is published under a CC-BY license.

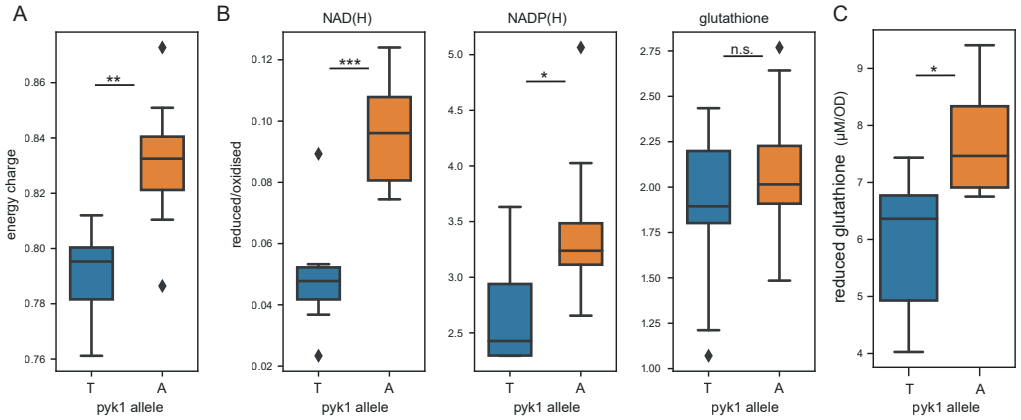


Fig. 2.6: Evidence for redox status differences in allele swap strain. **(A)** Energy charge of T- and A-strains calculated from concentrations of adenine nucleotides. **(B)** Ratio of reduced to oxidised redox co-factors, grouped by *pyk1* allele. **(C)** Abundance of reduced glutathione. Significance keys: *p $<$ 0.05, **p $<$ 0.005, ***p $<$ 0.0005 (Welch's t-test). All panels were reproduced with minor changes from Kamrad et al. (2020a) which is published under a CC-BY license.

metabolomics data, it is strong evidence that flux through Pyk1 is considerably lower in the T-strain compared to the A-strain.

The metabolomic data also enabled an assessment of cellular energy and redox status. Cellular energy charge, defined as

$$\mathrm{EC} = \frac{[\mathrm{ATP}] + \frac{1}{2}[\mathrm{ADP}]}{[\mathrm{ATP}] + [\mathrm{ADP}] + [\mathrm{AMP}]}$$

by Atkinson and Walton (1967), was 4.7% higher in the A-strain (Fig. 2.6A, p=0.002, Welch's t-test). Overall similar ranges have been reported for other organisms (De la Fuente et al., 2014). Reduced and oxidised isoforms were measured for NAD(H), NADP(H) and L-glutathione and we generally observed a higher ratio of reduced to oxidised forms (Fig. 2.6B). For NAD(H) and NADP(H) this difference was significant (p=0.0001 and p=0.024 respectively, Welch's t-test), but it was not for glutathione (p=0.289, Welch's t-test). Additionally, a significant increase in the concentration of reduced glutathione was detected in the A-strain (Fig. 2.6A, p_{adj}=0.015, Welch's t-test). These ranges of observed values agree with the paradigm that NAD(H) exists primarily in the oxidised form in order to provide electron acceptors to metabolic reactions, while the NADP(H) is predominantly reduced to serve as electron donor for anabolic reactions and the anti-oxidant response (Blacker and Duchen, 2016). The latter has been shown to be limiting when cells are subject

to oxidative stress as it provides reducing equivalents for the glutathione-, peroxiredoxin- and thioredoxin-dependent reduction systems (Vivancos et al., 2006; Drakulic et al., 2005; Veal et al., 2014; Carmel-Harel and Storz, 2000). Limitations of this experiment include the inability to distinguish between different sub-populations of the co-factors (Sun et al., 2012), eg located in different cellular compartments, and the danger of samples oxidising during sample preparation (Lu et al., 2018). So while the absolute quantification of redox states is possibly not exact, the relative differences are most likely real, as we made every effort to avoid systematic biases by randomising and blinding sample preparation and acquisition. Overall, this analysis suggests that the T-strain could be subject to more (presumably endogenous) oxidative stress than the A-strain.

2.2.3 Gene expression analysis reveals system-wide reconfiguration of growth programmes

In order to further characterise the cellular response to changed glycolytic flux, transcriptomes of both strains were characterised by RNA-seq. This analysis was done by Jan Grossbach, documented in detail in Kamrad et al. (2020a), and is only briefly summarised here. A total of 7750 transcripts, which included non-coding RNAs, were quantified for five biological repeats per strain. While *pyk1* expression levels were unaltered, a total of 960 transcripts were found to be differentially expressed at a false discovery rate (FDR) of $<10\%$ (Fig. 2.7A). Gene ontology (GO) terms significantly over-represented in the set of differentially expressed transcripts included mitochondrial respiration (down in the A-strain) and ribosome biogenesis (up in the A-strain) (Fig. 2.7B). The transcriptome response showed significant overlaps (in terms of sets of significantly affected genes) with that obtained during TORC1 inhibition (data from Rallis et al. (2013), $p=6.1\text{x}10^{-12}$, Fisher's exact test) and oxidative stress (data from Chen et al. (2003), $p=2.8\text{x}10^{-68}$, Fisher's exact test, Fig. 2.7C). The increased oxidative stress likely originates from higher respiration activity in the T-strain, a major source of endogenous reactive oxygen species, and which is coupled to an upregulation of cellular anti-oxidant systems (Grüning et al., 2011). Overall, these data indicate that the A-strain with the 'fixed' Pyk1 respires less and executes a cellular programme geared towards proliferation.

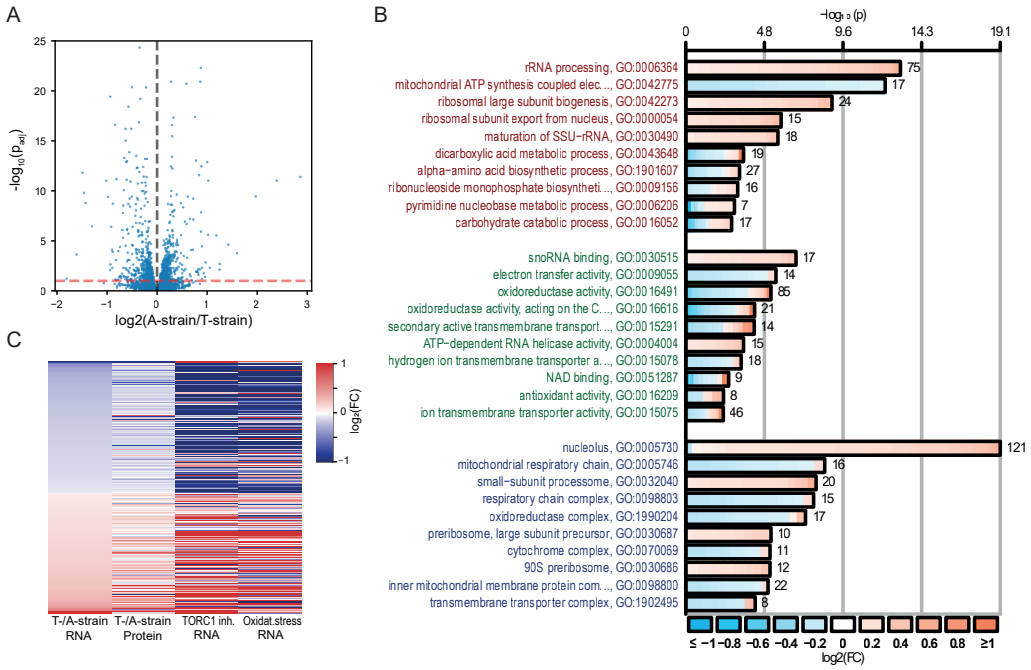


Fig. 2.7: Transcriptome analysis indicates physiological reconfiguration. **(A)** Volcano plot illustrating the results of differential expression analysis which identified 960 changed transcripts between T- and A-strain at an FDR of 0.1 (dashed line). **(B)** Cellplot visualising results of GO enrichment analysis on set of differentially expressed transcripts. Biological process (top), Molecular function (middle) and cellular component (bottom). The length of the bar represents the statistical significance of the enrichment while the colour within each bar describes the direction and magnitude of fold changes across all genes annotated to that term. **(C)** Comparison of transcriptome signature of allele-swap strains with published signatures of TORC1 inhibition and oxidative stress. Each (very thin) row corresponds to a gene and rows are ordered by fold-change in our expression dataset (first column). Gene expression at the protein level (second column) generally corresponds well with transcript levels, but there are some clear exception (please see Kamrad et al. (2020a) for a full discussion). The signature of the allele-swap strains is similar to that obtained for TORC1 inhibition (Rallis et al. (2013), third column) and oxidative stress (Chen et al. (2003), fourth column). This indicates that the T-strain, when compared to the A-strain, is in a cellular state of slower growth and higher stress.
All panels were reproduced with minor changes from Kamrad et al. (2020a) which is published under a CC-BY license. Panels B and C were originally created by Jan Grossbach.

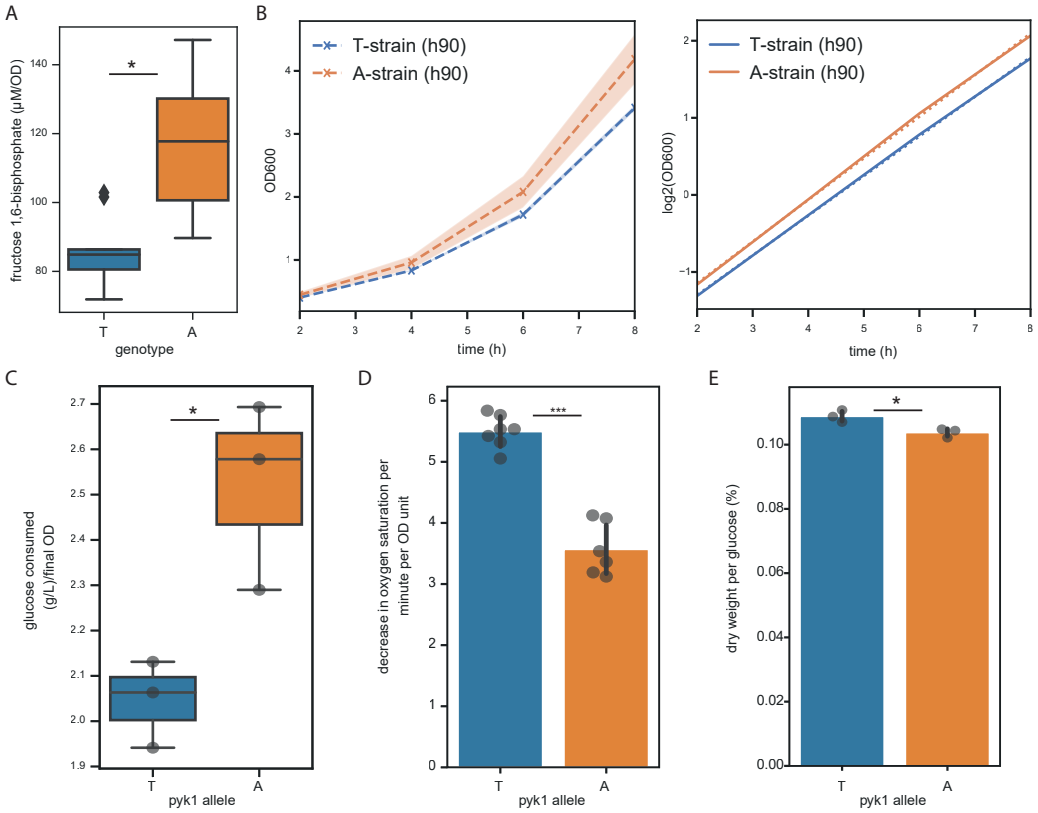


Fig. 2.8: Higher glycolytic flux results in faster growth and less respiration. **(A)** The A-strain has higher levels of fructose 1,6-bisphosphate, the levels of which are positively correlated with glycolytic flux. This data is part of the metabolomic dataset (Fig. 2.4; n=8 for A-strain, n=9 for T-strain). **(B)** Growth curve on linear (left) and logarithmic scale (right) based on OD measurements at 600nm. Three biological repeats of each strain where inoculated at the same OD and grown for eight hours. **(C)** Glucose uptake over 8 hours of growth (same cultures as in B) is significantly higher in the A-strain, even when corrected to final biomass. Glucose concentrations in the media at the beginning and end of the experiment were determined using a colourimetric assay kit. **(D)** Respiration rate is lower in the A-strain. The rate of decrease in oxygen saturation in exponentially growing cultures in rich YES media was measured with an amperometric probe. Shown are OD-corrected rates for 3 biological replicates of each strain, measured in technical duplicates. **(E)** Biomass yield (final dry biomass divided by the mass of glucose per volume of media) is significantly higher in the T-strain (3 biological replicates each).
Significance keys: *$p < 0.05$, **$p < 0.005$, ***$p < 0.0005$ (Welch's t-test). All panels were reproduced with minor changes from Kamrad et al. (2020a) which is published under a CC-BY license.

2.2.4 Higher glycolytic flux in the A-strain enables faster growth with less respiration

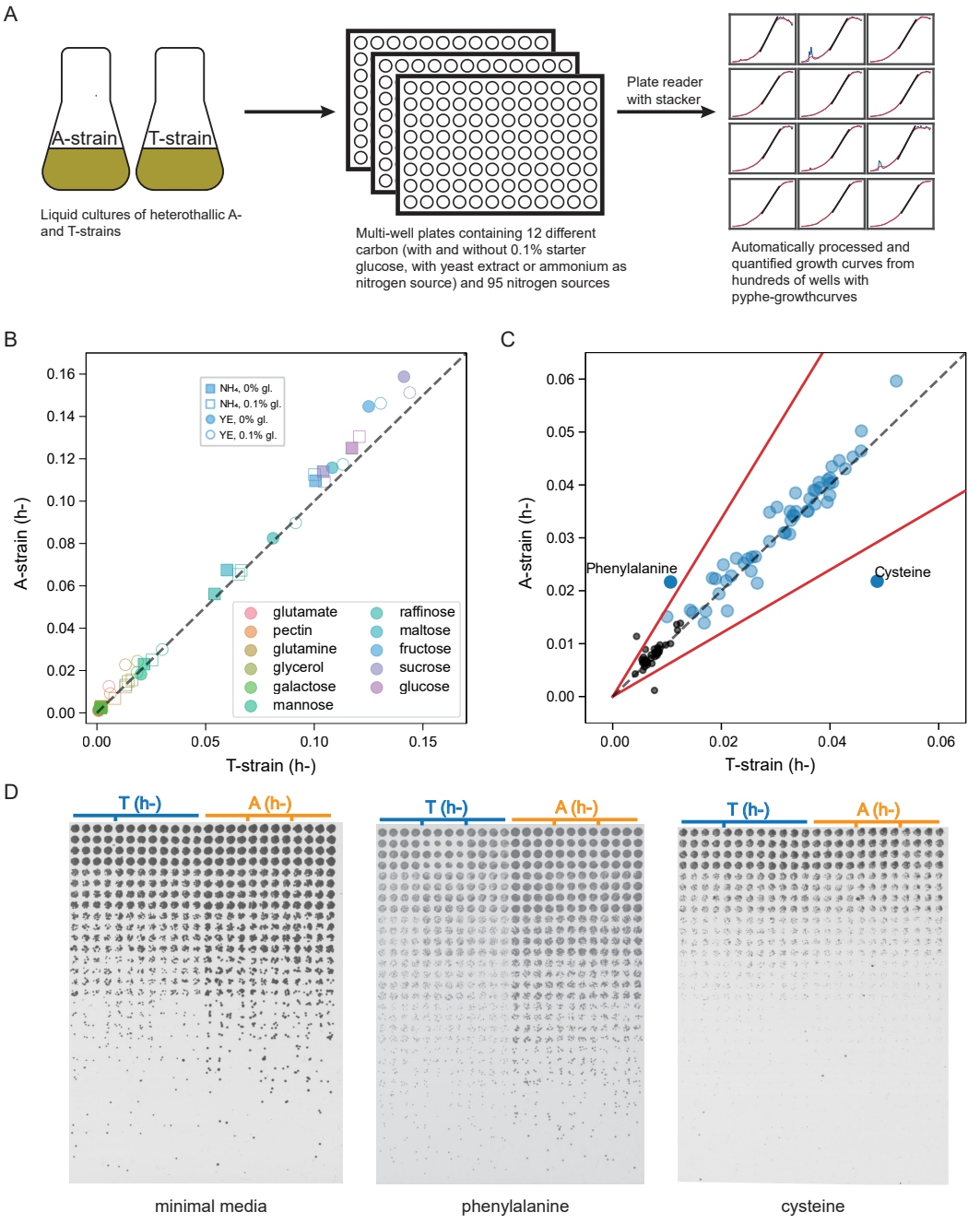
Another key finding from the metabolomic data included increased levels of fructose 1,6-bisphosphate (F16BP) in the A-strain (fold change=1.35, p_{adj}=0.015, Fig. 2.8A). F16BP is an example of a flux-signalling metabolite (Litsios et al., 2018) and its concentration is known to reflect glycolytic flux in several yeast species (Christen and Sauer, 2011; Huberts et al., 2012). Together with the gene expression data, specifically the upregulation of ribosome biogenesis, this strongly suggests that the A-strain grows faster, ie has a lower population doubling time. This was confirmed by measuring biomass (by OD600) every two hours over six hours of exponential growth for three biological replicates of each strain (Fig. 2.8B). This data was well described by an exponential doubling pattern (linear on a log2-scale) and doubling times were determined by fitting lines and taking the inverse of the slope. The A-strain showed a 4.7% decrease in doubling time: 1.85h down from 1.94h. The value for the standard lab strain (T-strain) was already somewhat lower than what has been previously reported by Petersen and Russell (2016) in the same conditions (2h 10min), but this is in the expected range of variation among laboratories. I furthermore confirmed the higher glycolytic flux in the A-strain with colourimetric glucose concentration assays on media at the beginning and after eight hours of growth. The A-strain showed significantly higher depletion of the media, ie higher glucose uptake, even when normalised to the final OD (Fig. 2.8C).

The gene expression data indicates reduced mitochondrial respiration activity in the A-strain and I confirmed this with oxygen uptake rate measurements using an amperometric probe (see Materials and Methods). The A-strain showed indeed a 35% lower rate in the OD-corrected decrease in oxygen saturation of the media over time (Fig. 2.8D, ratio of means = 0.65, p = 0.0002, Welch's t-test, three biological replicates, each measured in technical duplicates). Energy production by mitochondrial respiration is far more carbon efficient then energy production by fermentation and it is therefore probable that the T-strain will show a higher growth yield when the media is exhausted of glucose. This was tested by determining dry biomass of cultures after 24h of growth which confirmed a higher final biomass of the T-strain. The biomass conversion yield (final dry biomass divided by the mass of glucose per volume of media) was 10.38%±0.14% for the A-strain and with 10.88%±0.19% for the T-strain (p=0.02, Welch's t-test, three biological replicates each,

Fig. 2.8E). Overall, these data establish that the unusual T-allele found in the standard lab strain has a strong inhibitory effect on glycolytic flux and growth in rich YES media at 32°C, the condition used most widely by *S. pombe* researchers. Cells respond to this flux limitation by upregulating mitochondrial respiration which slightly increases carbon efficiency and final biomass.

So why is the T-allele found and maintained in a significant fraction of wild strains? Could there be an adaptive role that is not captured by growth in optimal, nutrient-rich laboratory conditions? A possibility is that the T-allele provides a growth advantage on some other carbon and nitrogen sources. I first estimated growth rates on 12 different carbon sources in 4 different backgrounds: with ammonium or yeast extract as nitrogen source, and with or without an additional 0.1% glucose. Such a small amount of 'priming' glucose is usually included for *S. pombe* when working with non-preferred carbon sources in order to buffer the transition shock between the glucose pre-culture and the main culture. This resulted in 48 unique carbon source regimes, in which each strain was tested in quadruplicates. The heterothallic strains with *h-* mating type were used for these experiments which cannot mate as they normally would in some of the nutrient poor conditions. Growth curves were recorded in a plate reader and maximum slopes extracted with *pyphe-growthcurves* (Kamrad et al., 2020b) (Fig. 2.9A). Glucose, *S. pombe's* preferred carbon source, supported the fastest growth for both strains and, as seen before in the precise, low-throughput growth experiments (Fig. 2.8B), the A-strain was found to grow faster. When ammonium is provided as nitrogen source instead of rich yeast extract, both strains grew slower (as expected) and the growth difference decreased. In minimal media, cells have increased biosynthetic requirements. Specifically, the synthesis of several amino acids requires TCA cycle intermediates and thereby active respiration (Malecki et al., 2020). So presumably, the slower growth makes the upper limit to glycolytic flux imposed by the T-mutation less relevant and both strains respire to some degree. On non-preferred carbon sources both strains had much lower growth rates and there was no observable difference between the strains.

Using a similar approach, I tested for growth differences on a wide range of nitrogen sources, this time relying on commercial assay plates with pre-loaded compounds (Biolog PhenotypeArrays) (Fig. 2.9C). The 95 nitrogen sources tested included amino acids, sugar-amines, di-peptides and various other small organic compounds. Strains were measured in a single replicate, making the data somewhat less reliable than for the carbon source


A
A-strain
T-strain
Liquid cultures of heterothallic A- and T-strains
Plate reader with stacker
Multi-well plates containing 12 different carbon (with and without 0.1% starter glucose, with yeast extract or ammonium as nitrogen source) and 95 nitrogen sources
Automatically processed and quantified growth curves from hundreds of wells with pyphe-growthcurves
B
NH4, 0% gl.
NH4, 0.1% gl.
YE, 0% gl.
YE, 0.1% gl.
glutamate
pectin
glutamine
glycerol
galactose
mannose
raffinose
maltose
fructose
sucrose
glucose
A-strain (h-)
T-strain (h-)
C
Phenylalanine
Cysteine
A-strain (h-)
T-strain (h-)
D
T (h-)
A (h-)
minimal media
phenylalanine
cysteine

screen. Two conditions showed a large difference between the two strains, with the A-strain growing better on L-phenylalanine as sole nitrogen source and the T-strain growing better on L-cysteine. Both hits were followed up with spot assays which confirmed the growth difference on phenylalanine (Fig. 2.9D). The hit on cysteine could not be validated and therefore is almost certainly a false positive. The scope of these two datasets was as wide as feasible, even though of course not all possible nutrients and combinations thereof could be tested. While containing largely negative results, these data importantly suggest that the T-allele does not provide an adaptive advantage for growth on specific nutrient sources, as would be plausible for a mutation in a key glycolytic enzyme. This raises the question if the *pyk1* locus affects other cellular traits, not in the closest sense related to metabolism.

2.2.5 The *pyk1* locus affects resistance to various stresses

Using spot assays, I first confirmed the initial GWAS hit and showed that the A-strain is more resistant to antimycin A (Fig. 2.10). In a complementary system in budding yeast, it has shown that reduced PYK activity and increased respiration triggers an upregulation of the anti-oxidant response via the PPP (Grüning et al., 2011). In line with this -and the

Fig. 2.9 *(preceding page)*: Growth of T- and A-strains on different carbon and nitrogen sources. **(A)** Overview of experimental workflow. Liquid cultures of both strains were grown in flasks to mid-exponential phase and then diluted into multi-well plates containing different carbon sources, with either ammonium or yeast extract as nitrogen source, and with and without 0.1% priming glucose. Growth was monitored by OD600 measurements in a plate reader and the obtained growth curves were analysed with *pyphe*. **(B)** Maximum slopes of growth curves of the T- and A-strain on 12 different carbon sources (colours of markers), each in 4 different media backgrounds (marker shape and fill). 4 biological replicates were measured in each condition and the mean is shown. No strong differences between the strains are apparent and in no condition does the T-strain outperform the A-strain. **(C)** Maximum slopes of growth curves of the T- and A-strain on 95 different nitrogen sources, based on a single biological replicate. Red lines show the chosen cut-off of abs(log2(A-strain/T-strain)) > 0.75. Conditions with a maximum slope below 0.015 in both strains were discarded. This indicates that strains differ in their ability to utlise phenylalanine and cysteine. **(D)** Spot tests for validation of results from (C). Three biological replicates of each strain were grown to mid-exponential phase and diluted to OD 0.4. 3-fold serial dilutions of each strain were prepared across rows of a 96 multi-well plate. Dilutions were then spotted 16 times into 1536 format, on agar EMM agar containing ammonium, phenylalanine or cysteine as sole nitrogen source. These assays could replicate the growth difference on phenylalanine but not on cysteine, suggesting that the latter was a false positive in the high-throughput screen.
Panels B+C and parts of panel D were reproduced with minor changes from Kamrad et al. (2020a) which is published under a CC-BY license.

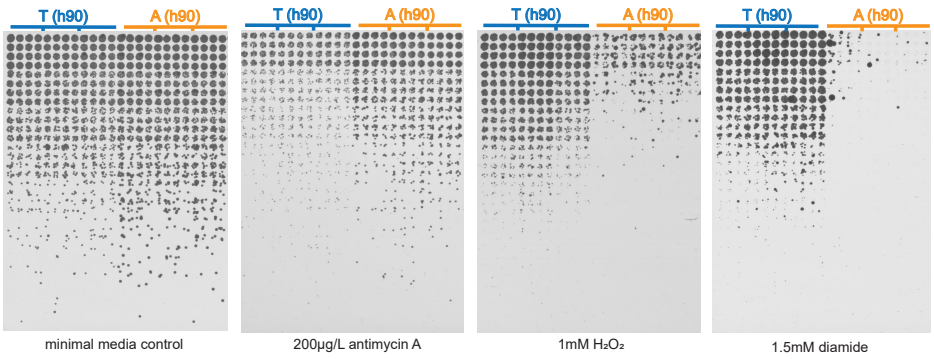


Fig. 2.10: The A-strain has decreased resistance to oxidative stress. Spot assays of serial dilutions of three biological replicates of each strain on media with antimycin A and two agents causing oxidative stress. A representative image is shown as control.
This figure was reproduced with minor changes from Kamrad et al. (2020a) which is published under a CC-BY license.

transcriptome data- the T-strain exhibited increased resistance to oxidative stress agents (hydrogen peroxide and diamide) in the media (Fig. 2.10). Such trade-offs between growth and stress resistance are a key concept in evolutionary adaptation (Ferenci, 2016), and we wondered if the A-strain might show altered resistance to stresses not previously associated with PYK activity.

In order to assess the wider impact on stress resistance, we screened 72 drugs and toxins for differential resistance of the T- and A-strain. Biolog PhenotypeArrays, which contain four different concentrations of each compound pre-loaded in each well, were used to obtain growth curves in high-throughput for a single replicate of both strains (similar set-up as Fig. 2.9A). Using an arbitrary cut-off of abs(log2(A-strain/T-strain))>0.75, nine compounds resulted in differential growth in total. The A-strain showed increased resistance to D-L-Alanine hydroxamate, barium chloride, chlorpromazine (which causes membrane stress, De Filippi et al. (2007)), caffeine (which has pleiotropic effects, including TOR inhibition, Rallis et al. (2013)), capreomycin (which binds to ribosomes, Lin et al. (2014)), phenylarsine oxide (an inhibitor of tyrosine phosphatases, Oustrin et al. (1995)) and thallium (I) acetate (which is highly toxic due to its similarity to potassium ions and binds to mammalian PYK with a stronger affinity, but weaker activating effect, than potassium Reuben and Kayne (1971); Kayne (1971)). The T-strain was relatively resistant to two compounds, EGTA (a chelator of bivalent cations, including Mg^{2+} which activates *S. cerevisiae* PYK, Rhodes et al. (1986)) and acriflavine (an antiseptic). Four of these compounds

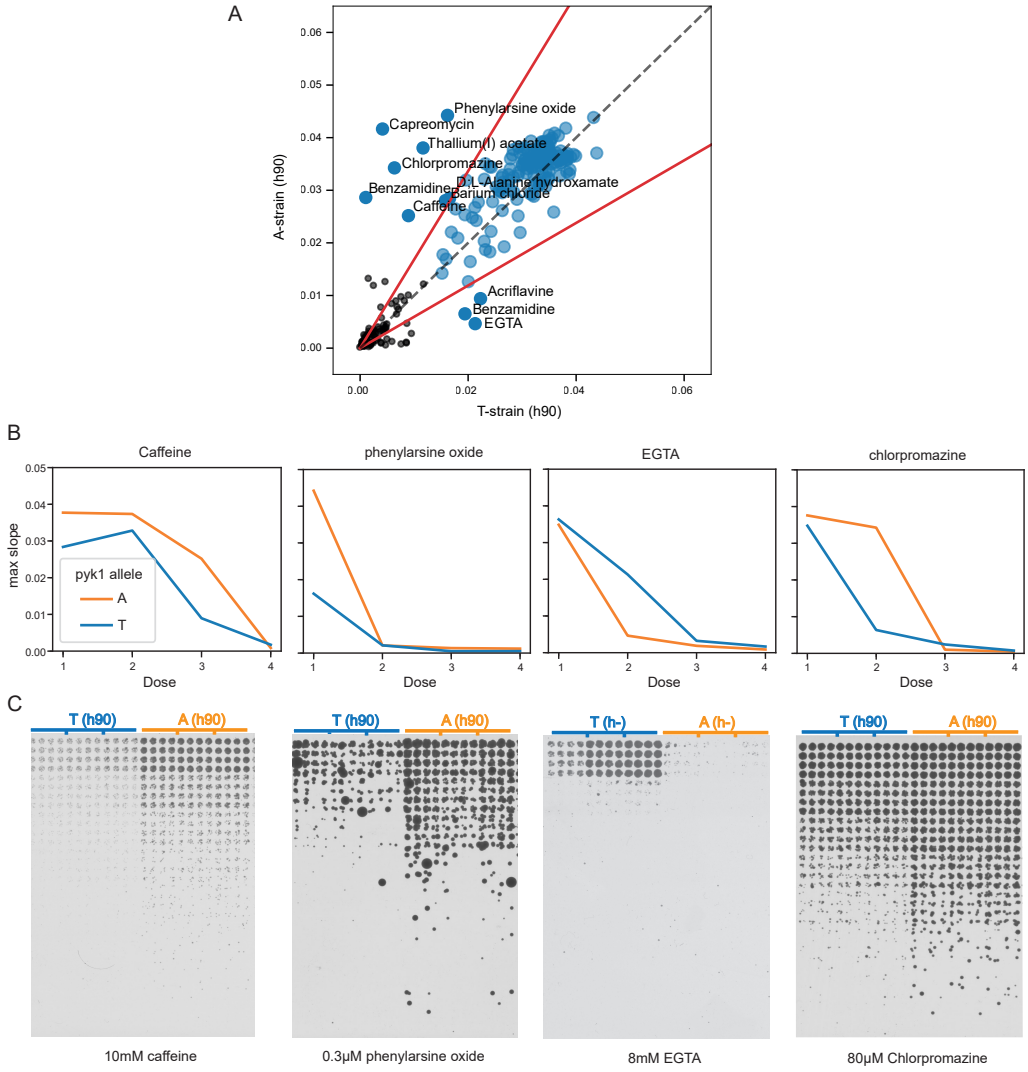


Fig. 2.11: Drug screen reveals wide-spread impact on stress resistance. **(A)** Maximum growth rates of T- and A-strain (single replicate per strain) on 72 drugs and toxins. Each compound was tested at 4 concentrations and all concentrations are plotted individually. For benzamidine, two doses were a hit and these hits were in opposite directions. **(B)** Selected dose response curves of compounds for which differential resistance was detected. The concentration of each dose is not available from the company. **(C)** Spot assays of serial dilutions of both strains (3 biological replicates per strain) confirming hits for all selected compounds. Control plates without toxins were included in each batch (a representative control is shown in Fig. 2.10). This figure was reproduced with minor changes from Kamrad et al. (2020a) which is published under a CC-BY license.

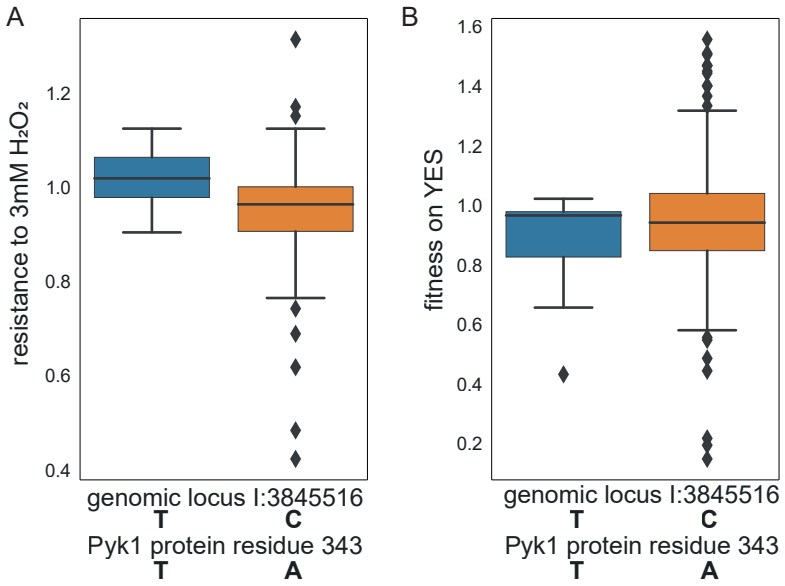


Fig. 2.12: Pyk1 affects oxidative stress resistance in the wild. **(A)** Resistance of 156 wild strains to 3mM hydrogen peroxide, grouped by *pyk1* allele. The T-strain features a higher mean resistance than A-strains (means and standard deviation 1.01±0.06 and 0.95±0.11, p=0.0021, Welch's t-test). **(A)** Fitness of 158 wild strains, grouped by *pyk1* allele, on standard rich YES media. No clear trend is apparent in the data and no significant difference was detected (means and standard deviation 0.88±0.16 and 0.94±0.24, p=0.18, Welch's t-test).
This figure was reproduced with minor changes from Kamrad et al. (2020a) which is published under a CC-BY license.

were selected based on their dose-response curves (Fig. 2.11B, all 72 curves can be found in Kamrad et al. (2020a)) and could be successfully validated with spot tests (Fig. 2.11C).

This set-up might initially appear highly artificial, with few of these compounds likely to be found in the wild. However, it specifically tests the cellular response to the (partial) inhibition of specific processes or cellular components, which may otherwise occur naturally through other mechanisms. Some of the conditions, specifically EGTA and thallium (I) in which differential resistance was detected can be directly related to Pyk1 function. But for the majority of hits, the mechanism is most likely more indirect. This suggests a general, pleiotropic role of glycolysis in modulating stress resistance, which is compatible with the wide-ranging impact of this mutation on the cellular gene expression programme (Fig. 2.7).

The impact of the mutation of interest has at this point been dissected in great detail on the molecular and phenotypic levels. But can findings in the laboratory background be generalised to wild populations? To address this question, I quantified oxidative stress resistance for all wild strains, using a colony screen similar to that outlined in Fig. 2.1A. The group of strains carrying the the T-allele was indeed significantly more resistant to hydrogen

peroxide (Fig. 2.12A, p=0.0021, Welch's t-test). This indicates that the *pyk1* locus really does affect oxidative stress resistance in the wild, giving support to the hypothesis that this is what selected the mutation. Basal growth on rich YES media on the other hand did not show a clear growth difference between strain groups (Fig. 2.12B, p=0.18, Welch's t-test). Two explanations are plausible here: either growth rate is a more complex phenotype, controlled by numerous loci which 'dilutes out' the effect of *pyk1*, or T-strains have counter-acting, compensatory mutations that have been specifically selected for.

2.3 Discussion

Using an allele swap in a tightly controlled genetic background, we have dissected the effect of a natural SNP in the *S. pombe pyk1* gene in detail. The minor T-allele limits glycolytic flux in standard rich media and this has a wide-ranging impact on gene expression, growth and stress resistance. This work recapitulates earlier findings which have shown that altered PYK flux changes the fermentation/respiration balance (Grüning et al., 2011; Pearce et al., 2001), but adds several important novel insights. First, while previous findings were obtained with *S. cerevisiae*, this work uses an evolutionarily distant, complementary model yeast species. This underlines the universality of the regulatory role of PYK. The *S. pombe* community only uses a single set of closely related reference strains (differing mainly in their mating type) and these strains are part of the minority of wild strains carrying the flux-limiting T-allele. This makes these strains unusual within their own species, but also across eukaryotes, with respect to PYK activity, which should be considered in future (metabolic) investigations of this important model organism.

Second, previous studies have used synthetic systems which have shown that PYK flux can regulate the respiration/fermentation balance and oxidative stress resistance, but these new data show that it actually does so in the wild. *S. pombe* is unusual in having only a single PYK isoform; most eukaryotes alternatively express low- and high-activity isoforms under specific conditions or cell types (Muñoz and Ponce, 2003; Allert et al., 1991; Israelsen and Vander Heiden, 2015; Bradley et al., 2019; Bluemlein et al., 2011). This makes *S. pombe* an ideal, simple model system for studying the impact of changes of PYK flux, as wild type strains are permanently locked into either the high- or low-activity state. This, together with the distribution of the allele in the wild in at least two different lineages and the associated phenotypes, suggest an adaptive role for the T-allele. Our extensive screens give no indication that this was an adaptation to growth on different carbon or nitrogen

sources, as could perhaps be expected for a variant in a glycolytic enzyme. Instead, the data point towards the known and conserved link to stress resistance, in particular oxidative stress (Shimizu and Matsuoka, 2019). This is ecologically relevant in niches where yeasts are known to form microbial communities with hydrogen peroxide-producing lactic acid bacteria (Ito et al., 2003; Ponomarova et al., 2017). More generally, I believe this is an example of a poorly explored wider phenomenon in which metabolism governs seemingly unrelated cellular traits. Uncovering those connections, systematically, not anecdotally as done here, would be a promising avenue of research. A possible first step would be a comprehensive characterisation of a segregant library (Liti and Louis, 2012; Cubillos et al., 2013) with respects to metabolism and stress resistance, looking for genomic variants linked to both.

Third, the impact of changed PYK flux has been characterised for the first time on the systems level, revealing a wide-ranging reconfiguration of the cellular metabolic gene expression programme. We find a distinct signature associated with fast growth (in particular upregulation of ribosome biogenesis) in the A-strain. It is often unclear and a matter of some debate if such signatures are the cause or consequence of faster growth, or in other words if metabolism regulates gene expression or vice versa. By delicately altering PYK flux by allele swapping, our data shows that increased flux of an intracellular metabolic reaction can be the singular and original cause of such wide-ranging gene expression changes. In particular as *S. pombe* misses multiple PYK isoforms and is therefore unlikely to have distinct, evolved gene expression responses to low and high PYK activity. This is an example of an emerging principle, which describes cells having distinct modes or programmes triggered by solely intra-cellular cues. Another example include the Crabtree-negative yeast *Komagataella phaffii* (*Pichia pastoris*) which can be made Crabtree-positive by overexpression of a single transcription factor (Ata et al., 2018). This likewise seems to me like a promising area of further investigation but for a deeper understanding it is crucial to systematically investigate what sort of intracellular events can cause such transitions.

So what exactly is the mechanism by which altered PYK flux results in these sweeping gene expression changes? It has been proposed that cells are able to 'sense' metabolic fluxes using flux-sensing metabolites (Litsios et al., 2018). There has been robust evidence in multiple species that fructose 1,6-bisphosphate (F16BP) is such a metabolite, responding specifically to glycolytic flux (Christen and Sauer, 2011; Huberts et al., 2012) and this is in

agreement with the increased F16BP levels observed in the A-strain. In the past, it has been proposed that a flux-sensing mechanism could regulate the balance between respiration and fermentation (Huberts et al., 2012) and this work provides some indirect evidence in support of this theory. The flux-sensing metabolites could then interact with transcription factors responsible for orchestrating responses (Wellen et al., 2010). This is a promising area of future investigation, where powerful concepts are in need of experimental investigation and validation.

2.4 Materials and Methods

A comprehensive description of the Materials and Methods can be found in Kamrad et al. (2020a) and the references therein. Selected methods for experiments I performed are elaborated on below.

Wild strain phenotyping and GWAS. Phenotyping of wild strains (Fig. 2.1) was done with the *pyphe* pipeline (Kamrad et al. (2020b), also described in detail in Chapter 1). 159 strains were used in these experiments, which is two fewer than in the original library which comprised 161 (Jeffares et al., 2015), due to contamination of the cryostocks. Using a programmable robotic colony picker (Stinger attachment for the RoToR HDA, Singer Instruments), strains were arranged in 384 format around a single grid of 96 control colonies of the standard lab strain 972. Randomly interspersed replicates of the standard lab strain were used to monitor technical noise. In retrospect, using the 1536 format would have been the better choice, allowing more replicates and denser placement of controls, but these experiments were done very early on during the development of *pyphe* and the data were of sufficient quality for the needs of the project. The arranged plates were passaged once on rich YES media (incubated for two days at 32°C) and then copied onto fresh YES plates with and without antimycin A and hydrogen peroxide. After incubation for two days, plates were scanned by transmission scanning (Epson V800 Photo) using *pyphe-scan*, colony sizes were determined with *gitter* (Wagih and Parts, 2014) and corrected fitness estimates were obtained by grid correction (Zackrisson et al., 2016) as implemented in *pyphe-analyse*. Quality control steps included removing very small colonies (smaller than 10 pixels at 600dpi resolution, which are most likely due to pinning errors), removing colonies with abnormal circularities reported by *gitter* (>1.1 or <0.85) and removing strains for which the phenotypic variance was greater than the variance of the entire data set after correction.

In preparation for GWAS, phenotype values were transformed to have a normal shape

(using the Box-Cox method), a mean of zero and unit variance using PowerTransformer from the *scikit-learn* package (Pedregosa et al., 2011). Small genomic variants were re-called from the published sequence data (Jeffares et al., 2015) using *freebayes* (Garrison and Marth, 2012). This re-calling of variants, instead of using those published in Jeffares et al. (2015), was mainly motivated by the fact that the published variants contained separate files for single nucleotide polymorphisms (SNPs) and insertion/deletions (indels), as they were being called separately using now out-dated tools. The two variant call sets are not trivial to merge/combine, whereas the modern haplotype-based caller *freebayes* detects SNPs, indels and other types of variants together (Garrison and Marth, 2012). The following options were passed to *freebayes*: –ploidy 1 –standard-filters –min-coverage 10 –min-alternate-count 3. The reference genome used was ASM294v2. Stringent quality control was applied to variants, including removing SNPs within 3bp of indels (using the –SnpGap option of *bcftools* Li et al. (2009)), removing low quality calls with a score below 30, removing variants not called in more than half the population and removing variants with a minor allele frequency below 5%. *SnpEff* (Cingolani et al., 2012) was used to predict variant impacts.

GWAS and heritability analysis (Fig. 2.2) was performed in *LDAK5* (Speed et al., 2012, 2017) using default parameters throughout. Doug Speed kindly advised on these analyses. For constructing the kinship matrix, predictors were first cut, thinned and weighted and the matrix was calculated using the direct method. Restricted maximum likelihood (REML) estimation as implemented in *LDAK5* was used for heritability estimation. A linear mixed model was used for associating individual variants with phenotypes, using the kinship matrix to correct for population structure.

The phylogenetic tree in Fig. 2.3 is based on biallelic SNPs in the region ±500bp up- and downstream of the *pyk1* gene (I:1:3844243-3847145) which were extracted using *bcftools* (Li et al., 2009). A pseudo alignment was generated with *VCF-kit* (Cook and Andersen, 2017) and a tree was constructed using the neighbour-joining method as implemented in *ClustalW2* (Larkin et al., 2007). The tree was drawn in *seaview* (Gouy et al., 2010).

Enzyme assays. PYK activity (Fig. 2.5) was assayed as previously described (Gehrig et al., 2017) with additional advise and reagents provided by Jamie Macpherson from the Anastasiou group at the Francis Crick Institute. Activities were determined for fresh (not previ-

ously frozen) lysate, but not purified Pyk1 protein which would have required a substantial effort. Exponentially growing cells were collected by centrifugation (800g, 3 min, RT) and broken by mechanical lysis with glass beads in a pH7-buffered solution containing 10 mM Tris at pH 7, 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT. DTT was included as a reducing agent to prevent oxidation of Pyk1 which risks inactivating it (Anastasiou et al., 2011). All reagents and lysate were kept on ice to prevent degradation. 96 multi-well plates were used to monitor reactions, where each well contained 200µl of aqueous 10 mM Tris at pH 7, 100 mM KCl, 5 mM $MgCl_2$, 20 µg L-lactate dehydrogenase from rabbit muscle, 5 mM ADP (all by Sigma-Aldrich), 10 mM PEP (Molekula) and 200µM NADH (Bioworld). The reaction mix was pre-incubated in the plate for two minutes at 32°C and the reaction was started by addition of the lysate. The NADH concentration was monitored by absorption measurements at 340nm (Tecan Infinite M200 Pro). The molar extinction coefficient of NADH (6220 M^{-1} cm^{-1}) and the path length (calculated from the reaction volume and the well diameter according to the manufacturer) were used to convert absorbance measurements to NADH concentrations. Initial values were found to be in very good agreement with the NADH concentration in the reaction mix. Lines were fitted to background-subtracted concentration traces using least squares regression. Slopes were corrected by the OD of the culture at the time of sampling which yielded a value proportional to PYK activity. I also tried correcting slopes by protein concentrations obtained by a multi-wavelength absorbance approach (Lunatic, Unchained Labs). This can theoretically correct for variations in extraction efficiency but was found to perform worse than correction by OD in practice.

Glucose uptake measurements. Direct confirmation of higher glucose uptake in the A-strain (Fig. 2.8C) required measurements of glucose in the media. A commercial, colourimetric kit (Glucose (HK) Assay Kit, Sigma-Aldrich, catalogue number GAHK20) was used for this purpose which is based on the enzymatic conversion of glucose to 6-phosphogluconate. This reaction reduces an equimolar amount of NAD^+ to NADH, which can then be quantified by absorbance at 340nm. I slightly modified the protocol of the manufacturer by reducing reaction volumes to 200µl and preparing a simplified standard curve with a serial 2-fold dilution. Samples were first diluted 1 to 20 in water and 10µl were added to 190µl of reaction mix prepared according to the manufacturer's instructions. After 30min incubation at room temperature in the dark, absorbance was measured (Tecan Infinite M200 Pro) and converted to concentrations using the standard curve.

Oxygen consumption rate measurements. Comparing the oxygen consumption rates of both strains (Fig. 2.8D) required a method for reliably quantifying dissolved oxygen in culture media. I initially attempted to do this in multi-well plates equipped with oxygen optodes (an optical sensor that can be read spectroscopically) using a microfermentor (BioLector, M2P labs). A plate with exponentially growing cultures was sealed air-tight using multiple layers of adhesive film and the oxygen concentrations recorded. I was unable to obtain reliable oxygen traces with this method, eg controls grown on antimycin or on the non-fermentable carbon source glycerol did not show the expected responses. We next considered using a Seahorse analyser (Agilent) but decided against this based on previous negative experiences using this analyser with yeast cells which are small and not attached. In the end, I used traditional, low throughput measurements with an oxygen meter. Exponential cultures were transferred into 25ml Erlenmeyer flasks and the probe (Hanna HI 98193) was inserted so that the flasks were completely filled. These were then taped shut with multiple layers of parafilm, making sure no air bubbles were trapped in the flask. This set-up ensures no oxygen enters or leaves the culture medium so that the rate of decrease in the oxygen concentration reflects oxygen consumed by cells alone. The oxygen saturation was recorded approximately every minute for 7min and lines were fitted to the traces.

Phenotypic screens. Phenotypic screens for growth on different toxins, carbon and nitrogen sources were performed in liquid culture in multi-well plates. For nitrogen sources and toxins, a commercial solution was used (Phenotype MicroArrays PM3, PM22, PM23 and PM25; Biolog) which provides plates with dry compounds pre-loaded in each well. Cultures were diluted to OD 0.05 in the appropriate media (EMM without ammonium chloride for nitrogen sources and YES for toxins) and pipetted directly on top of plates. Growth curves were recorded in a plate reader with stacker module (EnVision 2104, PerkinElmer) in a room that was not strictly temperature controlled but was stable at at 23.5±1°C over the course of the experiment. Growth curves displayed a substantial amount of noise and artefacts, such as drastic spikes or drops in the signal. This was most likely due to incomplete mixing of cultures in the well, a common problem with growth curves in 96 well format, even when mixing functionalities of the plate reader are used. This was compensated, as much as possible, by smoothing growth curves prior to further analysis using a median filter (size 5) and then a Gaussian filter (sigma=3). Maximum slopes were extracted with *pyphe-growthcurves* (Kamrad et al., 2020b) over a range of 12 timepoints. This wide window size

for slope extraction further minimises the impact of technical noise.

A carbon source Phenotype MicroArray is available from Biolog, containing 96 carbon sources, but this was found to be unsuitable for our needs in this case. *S. pombe* did not grow on the vast majority of these carbon sources and I wanted to further characterise the ones on which it does. I therefore compiled a list of carbon sources based on the literature and experience from our groups. Carbon sources were used at the same molarity as glucose in standard EMM media (2% w/w, 111 mM), except for the di- and tri-saccharides sucrose, maltose and raffinose where the concentration was adjusted to the number of hexose monomers. Lower concentrations were also used for pectin and glutamine due to low solubility. Cultures were prepared by combining carbon sources with nitrogen sources (yeast extract or ammonium chloride) and with/without additional priming glucose (final concentration 0.1%). Growth curves were recorded at 32°C in 384 well plates with a culture volume of 80µl (Tecan Infinite M200 Pro). This allowed for more replicates to be included but also slightly exacerbated the mixing problem. Growth were processed as described above.

2.5 Additional acknowledgements

This project was done in collaboration with the laboratory of Prof Andreas Beyer at the University of Cologne in Germany and the laboratory of Paola Picotti at ETH Zurich, Switzerland. Gorjan Stojanovski, an MRes student I was supervising at the time, helped with the experimental part of wild strain phenotyping. María Rodríguez-López designed and produced the plasmid for construction of allele replacement strains. María Rodríguez-López and StJohn Townsend constructed the allele replacement strains. María Rodríguez-López helped with sample collection for multi-omics experiments. Jan Grossbach, Valentina Capalletti, Maria Rodrigues-Lopez and Michael Mülleder prepared samples and acquired mutli-omics data. Jan Grossbach analysed gene expression data. Jan Grossbach produced an initial alignment of *pyk1* homologues, which was expanded by me. Profs. Jürg Bähler, Markus Ralser and Andreas Beyer supervised the study. Additionally, Jamie Macpherson advised on enzyme activity assays and Doug Speed advised on GWAS.

Chapter 3

Investigating amino acid uptake patterns by proteomics

It has been hypothesised that yeast grown on minimal media containing only ammonium as nitrogen source engage in cooperative exchange of amino acids. In this model, overflow results in the accumulation of amino acids in the extracellular environment which then triggers a heterogeneous response where some cells switch to being consumers of these compounds. This work investigates for the first time a key aspect of this model in detail: the response of *S. cerevisiae* cells to lowly concentrated extracellular amino acid supplements in the presence of high concentrations of glucose and ammonium. Using heavy carbon isotope (^{13}C) labelling, I find that a subset of the proteogenic amino acids are taken up with high affinity and suppress cell-internal synthesis completely. These amino acids tend to have a higher molecular weight and are non-preferred nitrogen source compared to the other set of amino acids which do not suppress their own synthesis. A novel, proteomics-based method is developed which can detect sub-populations differing in their amino acid uptake behaviour. Specifically, I hypothesise certain concentrations will result in some cells being consumers for it while other remain producers, in a binary switch-like manner. When applied to the standard laboratory strain BY4741, no evidence of such sub-populations is found. Nevertheless, I show that this conceptually novel method works in principle and that it could be used to deconvolute the proteome profiles of producer/consumer sub-populations. When applied to different strains and species, and in different growth conditions, this could open up a new avenue for investigations into single-cell metabolic heterogeneity.

3.1 Introduction

3.1.1 Overflow metabolism shapes microbial niches

When microbes grow on abundant and preferred nutrient sources, metabolites are exported and accumulate in the extracellular medium. A classical and well-studied example for this is fermentation (also known as aerobic glycolysis), a phenomenon where many yeasts suppress respiration even in the presence of oxygen when grown on preferred carbon sources like glucose. These Crabtree-positive species (Crabtree, 1929) export metabolic by-products (usually ethanol) in large quantities, which is referred to as overflow metabolism. The reason for this seemingly wasteful behaviour has been the focus of intense study and debate and has been discussed in more detail in Chapter 2. Suffice it to say that ethanol production and export has a profound impact on the surrounding environment and other microbes in it. Once glucose is depleted and cells have undergone the di-auxic shift, ethanol is used as a carbon source and enables a second growth phase. More recently, evidence has emerged that demonstrates the export of a much wider range of metabolites from growing cells, not typically considered waste products. Paczia et al. (2012) quantified metabolites in the extracellular medium (the exometabolome) and found many central carbon metabolism intermediates and amino acids in the micro-molar range during exponential growth on minimal glucose media containing ammonium as the sole nitrogen source. Campbell et al. (2015) showed that the extracellular fluid surrounding cells in *S. cerevisiae* colonies grown on minimal media contains amino acids, albeit at very low concentrations in the high picomolar range. These results are consistent with the more recent observation that yeast can support the growth of naturally auxotrophic lactic acid bacteria in minimal media, by the provision of amino acids (Ponomarova et al., 2017).

For metabolic intermediates and amino acids it is not immediately obvious why these valuable compounds, which have been produced at a cost, should be exported by cells. Broadly, three explanations have been proposed for this: export due to overflow (ie imbalances in the metabolic network that require secretions through active or passive mechanisms), accidental leakage (through membranes and imperfect transporters/channels) and cell death followed by lysis (Campbell et al., 2018). Cell death is negligible in typical lab cultures growing exponentially in standard glucose media and therefore unlikely to be the driver for the observed accumulation of extracellular metabolites (Paczia et al., 2012). And while accidental leakage cannot be excluded, there are good arguments for an ‘extended

overflow' model. Besides containing high amounts of glucose, standard media contain an excess of nitrogen, ie maximising fitness does not require an efficient utilisation of either carbon or nitrogen. Given the complexity and interdependencies of the metabolic network (>25% of metabolites are involved in >3 reactions), it might well be inevitable to overproduce certain amino acids in order to provide all biosynthetic building blocks at the required concentration (Campbell et al., 2018). Similarly, one could argue that, given a degree of biological noise and evolutionary imperfection, it is better to overproduce a metabolite rather than risking not making enough of it.

Specifically, the export of amino acids has been shown to increase upon TORC1 inhibition by either rapamycin or knocking-out transcription factors acting as TORC1 effectors (Ponomarova et al., 2017). Furthermore, the authors suggest that nitrogen catabolite repression (NCR), which inhibits the uptake and utilisation of non-preferred nitrogen sources in the presence of preferred ones, is involved in regulating amino acid export. The TORC1 effectors which increase lactic acid bacterial growth when knocked-out act as repressors of Gln3p and Dal81p (key activators of NCR-repressed genes such as permeases and catabolising enzymes). I find this result is somewhat counter-intuitive as an inactive NCR appears to promote export and is usually found in nutrient poor environments, when I would expect export to be higher in preferred nitrogen sources which enables lower doubling times. There is the possibility of this being a side effect of placing TORC1-inhibited cells in rich media, reducing growth, decreasing the amount of required amino acids and increasing the amount that is overflowing. Generally, much work remains to be done elucidating the purpose, mechanisms and regulation of metabolite export and more generally the numerous ways in which social microbes actively shape chemical and physical parameters of their niche and environment.

3.1.2 Yeast sense and respond to nutrient availability

The export of nutrients by *S. cerevisiae* can have obvious beneficial effects on other microbes, such as lactic acid bacteria (Ponomarova et al., 2017). But what effect do extracellular amino acids have on the yeast that export them? Yeast, like all free-living organisms, have evolved intricate mechanisms for sensing and responding to the availability of nutrients. But how exactly our standard lab strain will respond to the presence of lowly concentrated amino acids in the presence of the (highly concentrated) preferred nitrogen source ammonium is not known and will be at the core of this investigation. The following para-

graphs summarise the current knowledge of yeast nitrogen source sensing and regulation of their utilisation.

The *S. cerevisiae* genome encodes a range of amino acid permeases (AAPs). With few exceptions, the ones targeted to the plasma membrane belong to the APC (amino acid-polyamine-organocation) superfamily of secondary active transporters which use proton gradients and have 12 membranes-spanning segments (Bianchi et al., 2019). AAPs vary in their specificity and affinity. Gap1 and Agp1 are well described transporters with broad specificity, importing all or most amino acids. Several other AAPs transport multiple amino acids; eg Tat1 transports valine, leucine, isoleucine, tyrosine, tryptophan and histidine. Many amino acids have specific transporters, and often isoforms with high and low affinity (Zhang et al., 2018). The expression level and activity state of these is key to maintaining amino acid homeostasis in the cell. In nitrogen-poor conditions, cells express broad-specificity, high-affinity transporters, ie they are poised for amino acid uptake should they become available (Bianchi et al., 2019). On complex rich media (containing a wide range of nitrogen sources), cells consume different nitrogen sources at different rates and sequentially, with preferred nitrogen sources being consumed first (Crépin et al., 2012). *S. cerevisiae* cannot use cysteine, histidine or lysine as nitrogen source as it is lacking catabolic pathways for them (Bianchi et al., 2019). For several other amino acids (valine, leucine, isoleucine, methionine, and phenylalanine), the nitrogen can be assimilated but the remaining carbon skeletons (α-keto acids) are converted to 'fusel' alcohol or acids and secreted (Hazelwood et al., 2008) and these tend to be non-preferred and support lower growth rates. In the scenario of cells growing on nitrogen-poor media with amino acids becoming suddenly available, the challenge is therefore not primarily the uptake of these new nitrogen sources (for which transporters are already available) but rather selecting the desired nitrogen sources to be uptaken at the right concentration (eg by down-regulating general transporters and by expressing specific, lower affinity transporters). Accordingly, expressing a Gap1 isoform which cannot be ubiquitinated (Risinger et al., 2006) and overexpression of individual AAPs (Ruiz et al., 2017) results in amino acid supplements becoming toxic.

How amino acids are sensed and their import is regulated is an active area of research with many open questions. Several pathways that respond to the extracellular and intracellular nitrogen status of the cell have been described and they are briefly summarised here. Highly conserved are the TOR (target of rapamycin) and GAAC (general amino acid con-

trol) systems which sense the abundance (or lack) of intracellular amino acids (as well as other metabolites) and initiate concerted, pleiotropic responses affecting translation, gene expression, proliferation and autophagy (Efeyan et al., 2015; Conrad et al., 2014; Ljungdahl and Daignan-Fornier, 2012; Valvezan and Manning, 2019). Besides these universal systems, two yeast specific pathways have been studied in great detail. First, the SPS pathway is a key extracellular amino acid sensor in *S. cerevisiae*. The sensor protein, Ssy1, spans the plasma membrane and is homologous to amino acid permeases but has lost its transport activity. It forms a complex with Ptr3 and Ssy5. In the presence of extracellular amino acids, especially hydrophobic amino acids, Ssy5 cleaves inhibitory N-terminal domains off Stp1 and Stp2. Once activated, these become localised in the nucleus where they regulate the expression amino acid permeases (Ljungdahl, 2009). The SPS system, with its single sensor, is unlikely to be able to provide information about the exact identity of amino acids present. The second pathway, nitrogen catabolite repression (NCR) provides some of that specificity. In brief, it is the pathway with which the utilisation of non-preferred nitrogen sources is suppressed in the presence of preferred ones (Ljungdahl and Daignan-Fornier, 2012). It is not entirely clear how the NCR system senses specific amino acids, but it appears to be under TOR control and is unlikely to have external sensors (Zhang et al., 2018). Simultaneously, the strength of activation of NCR defines an amino acid as preferred on non-preferred, together with the growth rate it supports (Godard et al., 2007). One of the effects of NCR is the downregulation of the general permease Gap1, as well as other proteins facilitating the uptake and breakdown of non-preferred amino acids. Crucially, standard lab strains derived from S288c (such as the BY-series strains used in our laboratory) display a less strong NCR response in ammonium compared to strains derived from Σ1278b although it supports fast growth in both strains.

Only three individual AAP genes (Bap2, Uga4, and Ypq3) have been shown to be regulated by specific pathways (Bianchi et al., 2019), so a fine-grained, amino acid specific sensing system with compound-specific transcriptional responses appears to be largely lacking. A key question is how these signalling systems act together, specifically the relative importance/prevalence of measuring the extra- and intracellular nitrogen status. Experiments that I performed have recently shown that amino acids can accumulate intracellularly to many-fold higher levels compared to growth on minimal media containing only ammonium (Olin-Sandoval et al., 2019). This could initially proceed without a specific active

response of the cell and amino acids, once inside the cell, could act on intracellular sensors which appear to be more specific and intricate. Furthermore, the focus of most previous investigations has been the elucidation of 'classical' pathways ultimately regulating transcription. However, metabolism is to large parts regulated post-translationally, eg through allostery (Sauro, 2017). Increased concentrations of amino acids could directly act on the pathways that make it, eg through feedback inhibition (where a pathway end-product inhibits its own synthesis). Several recent studies have described how nitrogen signalling can regulate AAPs post-translationally. Particularly ubiquitin-dependent endocytosis, requiring the E3 ubiquitin ligase Rsp5, has been shown to likely be a key factor regulating AAPs, in turn regulated by TORC2 (Bianchi et al., 2019). Additionally, AAPs have been shown to be part of eiosomes (protein cluster domains on the plasma membrane), especially the MCC (microcompartment of Can1) domain. This has been suggested to reduce the likelihood of endocytosis but it is otherwise unclear if and how this affect transport function (Bianchi et al., 2019).

In summary, given the scenario of growth in minimal media with slowly accumulating amino acids, it is hard to predict how cells will respond, based on currently available knowledge. With an active NCR pathway, cells will be unlikely to start using these amino acids as nitrogen sources, but it is likely that they will be imported into cells, eliciting a metabolic response and being incorporated into proteins. Besides the mechanistic explanations discussed above, this would make sense evolutionarily, as importing and utilising available resources would reduce the burden on biosynthetic metabolism (Björkeroth et al., 2020). Many studies have indeed shown that prototrophic yeast do take up amino acid supplements in minimal (ammonium-rich) media: For histidine, uracil, leucine and methionine, our lab has shown that the prototroph wild type strain consumes these supplements at the same rate as derived strains auxotrophic for these compounds, giving some indication that the prototroph solely uses uptake to meet its biosynthetic requirements (Campbell et al., 2015). Fröhlich et al. (2013) have shown that supplemented lysine and arginine is incorporated into proteins while also repressing cell-endogenous synthesis. From these experiments and others, the paradigm emerged that yeast prefer to uptake an available metabolite from the environment rather than making it by itself (Campbell et al., 2018).

3.1.3 Clonal populations display single-cell heterogeneity

The two well-known phenomena described in the first two sections of this introduction, ie the observed accumulation of amino acids in the extracellular medium during growth and the ability and preference for uptake over synthesis, leads to an obvious paradox. If cells prefer uptake over synthesis, why are amino acids present in the media of yeast cultures? A proposed explanation is that cells in the culture respond heterogeneously to the presence of amino acids, resulting in the emergence of sub-populations. One population continues to produce amino acids while the other switches to consuming them (Campbell et al., 2018). Our lab has previously shown that yeast possesses the capacity for metabolite exchange at growth-relevant quantities using a segregating plasmid system which slowly introduces auxotrophic members into the population (Campbell et al., 2015). The resulting community is heterogeneous but grows only marginally slower than wild type cultures. However, if wild type prototrophic cells also differentiate into producer and consumer populations is entirely unclear and the focus of this chapter. The following section reviews existing evidence for heterogeneity in clonal *S. cerevisiae* populations, its triggers and downstream consequences.

One omnipresent source of heterogeneity is the fact that cells of proliferating cultures in nutrient-replete conditions are in different stages of the cell cycle, with distinct transcriptional programmes (Spellman et al., 1998; Saint et al., 2019). As mother cells repeat mitosis and budding, they grow in size and undergo replicative ageing, fundamentally altering physiology and metabolism (Leupold et al., 2019). Furthermore, nutrient depletion has been shown to illicit heterogeneous responses in multiple studies. First, entry into stationary phase has been shown to coincide with highly variable gene expression by single-cell RNAseq (Saint et al., 2019). Second, growth on media containing yeast extract and peptone at standard concentrations but low glucose (0.1%) has been shown to result in the emergence of two metabolically distinct populations. One is gluconeogenic and provides the disaccharide trehalose (made predominantly from aspartate) to the other population which is glycolytic and more proliferative (Varahan et al., 2019, 2020). Furthermore, when yeast colonies grown on solid agar media persist in a glucose-starved state, chronological ageing of the colonies results in populations of cells with distinct metabolic roles. Driven by ammonium export, colonies differentiate into an upper and lower layer of cells, with the upper layer retaining some degree of metabolic and proliferative activity while lower cells are starving, non-dividing and gluconeogenic (Palková and Váchová, 2006; Váchová and

Palková, 2018; Váchová et al., 2013).

In well mixed, nutrient-replete liquid cultures all cells experience the same environment and one would not expect the population to be heterogeneous due to extracellular factors. But still, individual cells differ from each other as recent studies have shown. Genome-wide libraries of GFP fusion proteins have been used in flow cytometry or microscopy screens to identify proteins whose expression is highly variable (Chong et al., 2015; Keren et al., 2015). On a smaller scale, the abundance of specific transcripts can be measured by fluorescent *in situ* hybridisation (Trcek et al., 2012). Only recently, single-cell RNA sequencing (scRNAseq) has been successfully applied to yeast cultures, capturing expression levels of a few hundred genes per cell (Saint et al., 2019; Nadal-Ribelles et al., 2019; Jariani et al., 2020). Each of these methods has advantages and drawbacks (eg the low genome coverage achieved with scRNAseq) which are beyond the scope of this introduction. But generally, a key finding is that metabolic genes can be highly variable between members of a population. For example, Nadal-Ribelles et al. (2019) identified several genes with highly variable expression annotated to sugar and amino acid metabolism, using strain BY4741 grown in rich complex YPD medium.

While most studies have focused on heterogeneity in the context of gene expression, a smaller number has explored metabolic heterogeneity. This is particularly challenging because the standard analytical methods in the field of metabolomics, namely chromatography coupled to mass spectrometry and nuclear magnetic resonance (NMR), are typically not sensitive enough for the investigation of single cells. Recently, some ultra-sensitive mass spectrometry imaging workflows are beginning to emerge, again with limits in either resolution, the type of observed compounds and the ability to investigate live cells (Kumar et al., 2020). Most single-cell metabolism studies to date have employed fluorescent sensors, usually limited to indicating a single metabolic parameter of the cell. Monteiro et al. (2019) have recently developed a biosensor able to quantify glycolytic flux in single cells. By the same group, the autofluorescent properties of NADH and a FRET-based ATP sensor have been used to investigate metabolic oscillations at the single-cell level and propose a link to the cell division cycle (Papagiannakis et al., 2017).

In summary, it is well established that cells in well-mixed clonal cultures differ from each other in terms of their life history, morphology, gene expression profile and metabolism. It has not been shown, but hypothesised, that clonal prototrophic cultures

differentiate into metabolically specialised sub-populations which exchange amino acids (Campbell et al., 2018). Besides the mechanistic considerations discussed in the previous sections, there are evolutionary arguments supporting this hypothesis. The metabolic network is inherently self-inhibitory (Alam et al., 2017). Division of labour could make metabolism more efficient for individual cells and the population and is widespread in microbial metabolism (Thommes et al., 2019; Zelezniak et al., 2015; de Lorenzo et al., 2015; Morris et al., 2012). This chapter presents a novel approach for detecting and characterising sub-populations which differ in their amino acid uptake behaviour. It is based on proteomics and ^{13}C labelling and identifies signatures of producer/consumer populations by investigating peptide labelling patterns. Results show that amino acids can be broadly classified into two groups: one which is taken up with high affinities and eventually replaces internal synthesis completely, and one which is only taken up at higher concentrations without ever completely suppressing self-synthesis. However, no evidence for co-existing consumer/producer populations is found. Rather, it is observed that supplemented amino acids are consumed until they are exhausted. This work provides new tools for investigating heterogeneity and suggests avenues for future applications.

3.2 Results

3.2.1 A peptide-based method for investigating sub-populations

The key requirement for metabolite exchange is the differentiation of clonemates into distinct sub-populations (producers and consumers), and for this, cells need to respond heterogeneously to the presence of extracellular metabolites. Presumably, this response is concentration dependent, ie the more of an amino acid is available, the more it will be taken up and utilised. This is easily rationalised for the border cases: without amino acid present, all cells have to be producers; with a large excess of an amino acid available, cells would be expected to switch from production to uptake of that amino acid for the reasons outlined in the introduction. However, the response between these two states, which occurs at intermediate concentrations of amino acid, is entirely uncharacterised. One would expect a gradual response of some kind, given the inherent complexity and noise of the underlying biological systems. A hypothetical dose-response curve illustrating this behaviour is shown in Fig. 3.1A. Following on from this, a key question is whether the response is binary (a switch) or gradual at the single-cell level, ie whether cells simultaneously import and produce amino acids. Both models, a gradual transition from synthesis to uptake

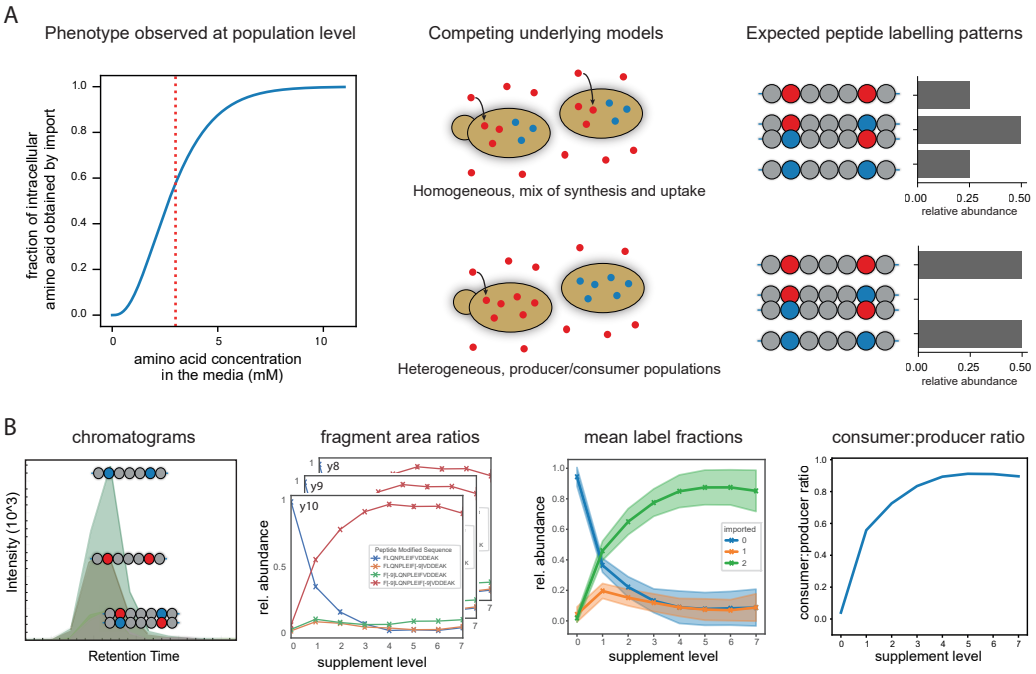


Fig. 3.1: ^{13}C-labelling patters in peptides can reveal sub-populations. **(A)** Left: Hypothetical dose-response curve for an amino acid supplied in the media. Without supplementation, the entire biosynthetic requirement is produced by cells. As the supplement concentration rises, it enters cells and a rising fraction of the required amount is imported rather than synthesised. Middle: Two competing models could explain the population phenotype observed on the left. Either cells respond homogeneously and switch gradually, or individual cells switch at different concentrations in a binary manner, resulting in producer/consumer sub-populations. Right: Shown are predicted peptide labelling patterns of a peptide containing two residues of the supplemented amino acid for both models. A binary switch results in the mixed labelling state (one of the two residues labelled) being absent (or rare). **(B)** Data analysis workflow for targeted peptide quantification by reverse-phase micro-flow liquid chromatography and tandem mass spectrometry. Cultures are grown in minimal media with ^{13}C glucose, ammonium and varying concentrations of a ^{12}C amino acid supplement. Produced and consumed amino acids therefore have different molecular masses and can be distinguished by mass spectrometry. From left to right: Suitable peptide fragment peaks are identified and integrated in *Skyline* (MacLean et al., 2010). Within each sample, the ratio of fragment abundances of the four different labelling states (two pure and two mixed states) is computed. Data for different fragments, peptides and biological replicates (usually n=2) are summarised. The relative abundance of peptides with one imported amino acid allow determining the switching mode. The global probability of incorporating ^{12}C (imported) amino acids is determined for each supplement concentration, resulting in a dose-response 'switching' curve.

homogeneously executed by the cells of the population or a heterogeneous binary switch, could explain the postulated population-level phenotype. I devised an experimental strategy to simultaneously characterise the overall fraction of produced/consumed molecules as well as the mode of switching at the single-cell level. Cells are grown in minimal media with ammonium as the sole nitrogen source and uniformly labelled ^{13}C glucose as the sole carbon source. Cultures are then supplemented with varying concentrations of unlabelled (^{12}C) amino acids (one at a time), which are distinguishable by mass from amino acid molecules synthesised by the cells. Cells are then harvested and proteins are extracted and digested with trypsin (Methods), resulting in ten-thousands of short peptides. Those peptides which contain the amino acid that was supplemented, have characteristic masses depending on the origin of the amino acid and can be quantified by liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS, Methods). Besides estimating the overall producer/consumer ratios, this method is informative about the mode of switching when peptides with two or more of the supplemented amino acid are considered. If cells react in a binary manner, ie are either full producers or full consumers, one would expect only fully labelled or unlabelled peptides. If the switch is gradual, one would expect the total number of labels in the protein to follow a binomial distribution (Fig. 3.1A). This 'pseudo single-cell' method therefore could enable the characterisation of sub-populations from an experiment that would typically only result in population-level averages.

Experimentally, this method built on existing internal expertise in proteomics workflows, especially that of Dr Christoph Messner. I was able to use his sample preparation protocol, liquid chromatography gradient and some mass spectrometry methods, all optimised for high-throughput yeast proteomics (Vowinckel et al. (2018); Messner et al. (2020), Messner et al. in preparation). Initially, a survey for suitable peptides (containing exactly two of each amino acid) was conducted by measuring the proteome of 96 cultures grown in minimal media supplemented with individual amino acids and supplement mixes, each in 4 biological replicates. Proteome profiles were acquired by sliding window acquisition (SWATH), a data independent acquisition strategy that simultaneously isolates and fragments groups of precursors falling in sequential mass windows. Peptides were identified and quantified from the fragment spectra using a spectral library (prepared by Christoph Messner) and our in-house analysis software DIA-NN (Demichev et al., 2020). Only peptides which were consistently identified across all runs, high in abundance and low in inter-

fering fragments were considered. After obtaining this list of suitable peptides, I developed targeted mass spectrometry methods to quantify the four different labelling states (none labelled, the first residue labelled only, the second residue labelled only, both residues labelled). Method development was done in *Skyline* (MacLean et al., 2010), creating a parallel reaction monitoring (PRM) method for each supplemented amino acid condition. In this acquisition mode, our Sciex QTOF mass spectrometer cycles through all precursor masses every couple of seconds, isolating the precursor using the first quadrupole, fragmenting it in a collision cell and recording the masses of the resulting fragments using time-of-flight (TOF). Mass spectrometry data is then inspected and different precursors and their labelling states are quantified in *Skyline*. Data for different fragments belonging to the same peptide, different peptides (8 were initially selected from the survey) and the biological replicates are then aggregated. The overall ratio of produced to consumed amino acid is determined from the overall proportion of sites occupied by labelled versus unlabelled amino acids, resulting in the final dose-response 'switching' curve (Fig. 3.1B).

3.2.2 Amino acids have widely differing uptake patterns

Using this method on batch cultures grown to OD 1.6, I recorded dose-response curves for 15 amino acids, at 7 concentrations and in biological duplicates each. A prototrophic derivative of strain BY4741, a very popular laboratory strain, was used throughout this project. Perhaps somewhat contrary to the paradigm 'Yeast prefer uptake over self-synthesis', incorporation patters differed a lot between individual amino acids (Fig. 3.2). Amino acids roughly fell into two groups which differed in two key characteristics. One group, which I call 'switcher' amino acids, are taken up with high affinities (defined by the critical supplement concentration required to achieve 50% labelling) and fully replace self-synthesis at higher concentrations. The other group, the 'non-switchers' (or strictly speaking 'late-switchers'), only showed significant incorporation at higher supplement concentrations and the supplement never fully substitutes for self-synthesis.

The partitioning of amino acids into switchers and non-switchers does not appear to be random. Non-switchers include key metabolites such as glutamate, glutamine and aspartate, which feed into numerous other pathways and act as nitrogen donors during the synthesis of other amino acids. Switchers on the other hand predominantly contain non-preferred amino acids and those degraded in the Ehrlich pathway. In order to quantitatively describe the response to supplementation, summarised switching curves were generated from the rela-

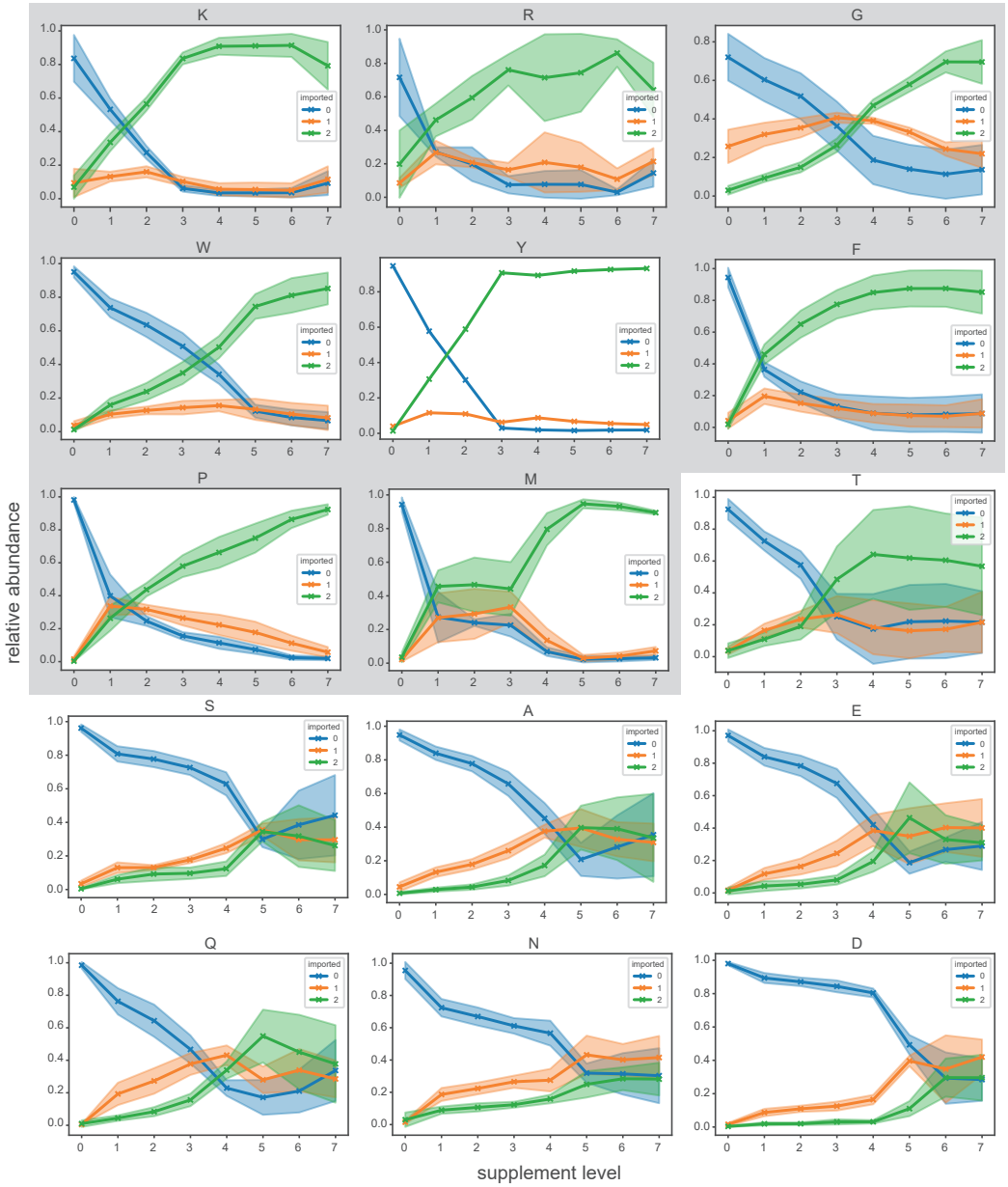
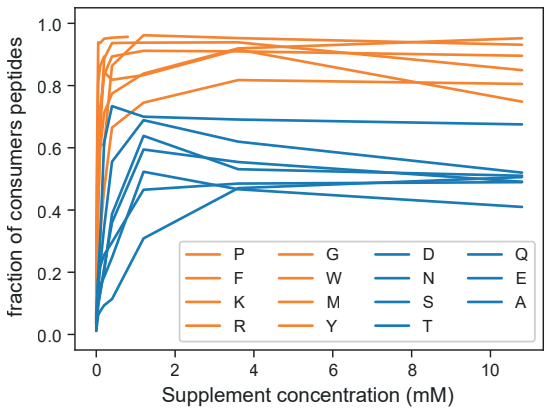


Fig. 3.2: Uptake patterns differ between amino acids. Dose-response curves showing the relative abundance of the three peptide labelling states (none, one or two imported ^{12}C amino acids) for cells grown with different individual amino acid supplements. Amino acids are denoted by the single letter code. 8 peptides were measured for each amino acid and measured in biological duplicates. Lines show the mean and shaded areas the standard deviation. Some peptides had to be excluded due to analytical problems (all but one in the case of tyrosine). Panels with a grey background switch at low concentrations and solely rely on uptake for higher supplement concentrations ('switcher' amino acids). The concentration levels (in increasing order are) 0, 0.05, 0.1, 0.2, 0.4, 1.2, 3.6, 10.8 mM, except in the case of tryptophan where they are 0, 0.003125, 0.00625, 0.0125, 0.025, 0.05, 0.1, 0.2 mM.

A



B

switchers		non-switchers	
a.a.	crit. c. (mM)	a.a.	crit. c. (mM)
K	0.068566	D	10.789315
R	0.0301	N	3.484402
G	0.226488	S	3.462492
W	0.004996	T	5.631521
M	0.033798	A	4.082556
Y	0.018143	E	0.760248
P	0.065268	Q	0.340851
F	0.041158		

Fig. 3.3: Switching amino acids have low critical concentrations and fully replace self-synthesis at higher doses. (A) Summary of switching curves from Fig. 3.2, showing the overall average proportion of labelled to unlabelled residues dependent on supplement concentration. Amino acids roughly fall into groups: switchers (orange) show a 50% incorporation rate at high µM concentrations (the critical concentration) and eventually reach almost 100% incorporation at higher concentrations, ie cells switch completely from self-synthesis to uptake. Non-switchers (blue) have higher critical concentrations and do not fully suppress self-synthesis. **(B)** Table listing critical concentrations for all amino acids tested, obtained by fitting cubic splines to the observed data.

tive abundance of different peptide labelling states (Fig. 3.2), describing the overall ratio of uptake to synthesis across all residues (Fig. 3.3A). Critical concentrations (at which 50% labelling is observed) were interpolated by cubic spline fits (Fig. 3.3B). I systematically explored between the physio(bio)chemical properties of the amino acids and their critical concentration (Fig. 3.4). Hydrophobic amino acids, which include the aromatic amino acids, tend to have lower critical concentrations, but switchers also notably include the positively charged amino acids arginine and lysine. In particular, the three aromatic amino acids (phenylalanine, tyrosine and tryptophan) had very low critical concentrations. This partly also drives the observed correlation with molecular weight and the number of carbons. In a first instance unrelated to these physical properties is the observed anti-correlation of critical concentrations with the abundance of amino acids in the proteome, with rare amino acids more likely to be switchers. Overall, these findings indicate a differentiate response which results in switching for less required and larger (and thus potentially more valuable) amino acids while maintaining the synthesis of crucial metabolic intermediates.

But how could the suppression of cell-internal synthesis be explained mechanistically? Using mass-spectrometry based metabolomics, I recently found out that many amino acids accumulate intracellularly to very high levels when present in the media, compared to un-

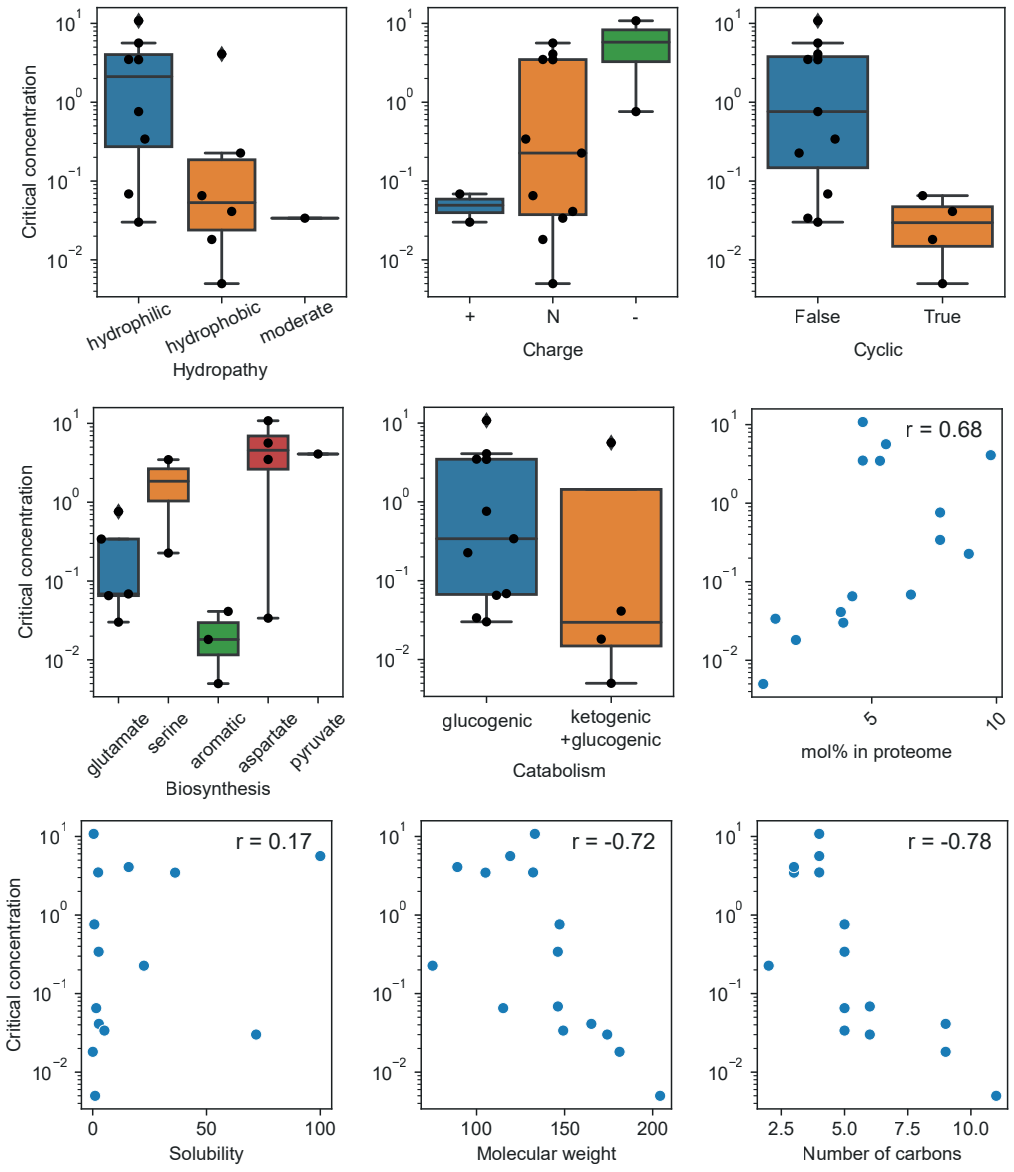


Fig. 3.4: Physio(bio)chemical properties vary between switchers and non-switchers. Panels show the critical concentration of amino acids from Fig. 3.3 on the y-axis and different potential explanatory properties on the x-axis. Hydropathy, molecular weight, number of carbons and charge data were obtained from Fisher. Data for 'Biosynthesis' were obtained from Rallis et al. (2020). The molar fraction of the proteome was obtained from Lange and Heijnen (2001). The Spearman rank correlation is shown in plot with continuous explanatory variables.

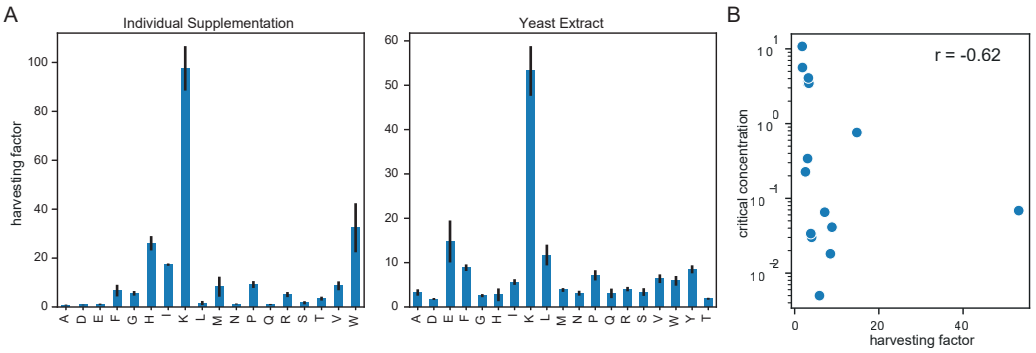


Fig. 3.5: Amino acid harvesting could suppress synthesis. (A) Wild type cultures were grown in minimal media with and without individual amino acids (0.334mM) and yeast extract (2%). Each condition was measured in biological quadruplicates. At mid-log phase, intracellular amino acid abundances were determined by LC-MS/MS as described by (Mülleder et al., 2016) and in Methods. Harvesting factors were determined as the ratio of abundance in supplemented versus unsupplemented media. Many amino acids, especially lysine accumulate to much higher levels when present in the media. The data underlying this panel was published in Olin-Sandoval et al. (2019). **(B)** This plot shows the critical concentration of amino acids from Fig. 3.3B on the y-axis and the harvesting factor for individual supplements from A on the x-axis. The r-value shown is the Spearman rank correlation between the two variables. This could mean that switcher amino acids inhibit cell-internal synthesis by accumulation and substrate inhibition.

supplemented minimal media. This phenomenon was termed harvesting and can be quantitatively described by the ratio of intracellular concentrations with versus without supplement (Olin-Sandoval et al., 2019) (Fig. 3.5A) . Interestingly, switcher amino acids tend to have higher harvesting factors (Fig. 3.5B). This could represent a simple, post-translational mechanism for suppressing synthesis by substrate inhibition. Enzymology and metabolic modelling could substantiate this hypothesis in future work.

3.2.3 Evidence for binary switching

Measuring peptides instead of free intracellular amino acid pools has the crucial advantage of potentially revealing the mode of switching at the single-cell level. If individual cells only ever exclusively produce or consume, one would expect peptides where one of the two residues is labelled to be rare. From the abundance of the three labelling states, the overall probability of incorporation p can be calculated:

$$p = 0a_0 + 1(a_1^1 + a_1^2) + 2a_2$$

where a_i are peptide abundances for peptides with i imported residues, and a_1^1 and a_1^2 are the two unique peptides with one labelled and one unlabelled reside. From this, the expected relative abundance b of labelling states, if p were constant and equal for all cells, can be

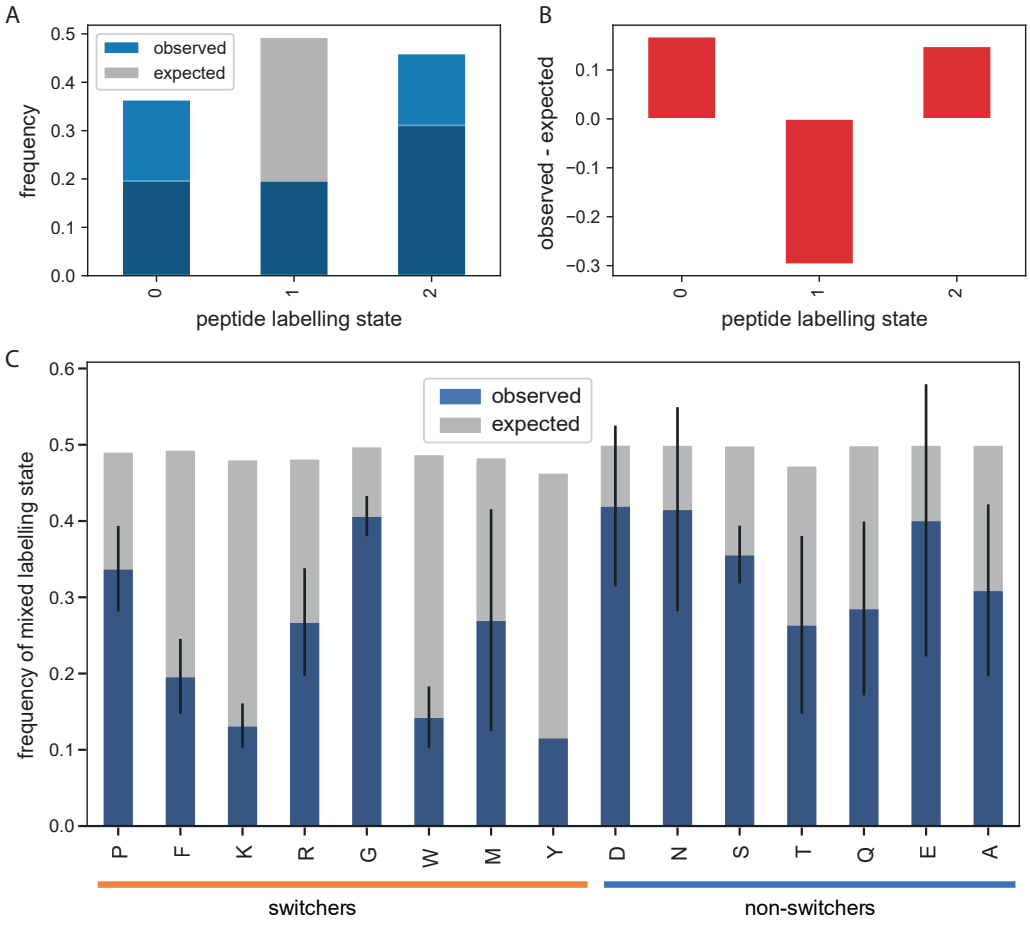


Fig. 3.6: Depletion of mixed labelling states suggests binary switch. **(A)** Average abundance of peptide labelling states for cells supplemented with 0.05mM phenylalanine (same data as shown in Fig. 3.2 for concentration level 1). The overall incorporation probability p was estimated as the average proportion of unlabelled (imported) residues and was 0.56 in this case. From this probability, the expected relative abundances of each labelling state were calculated based on the binomial probability distribution (see text). **(B)** Subtracting the expected frequencies from the observed ones, one notices that the mixed labelling state is less abundant than expected. This is evidence for a binary metabolic state, where cells either consumer or produce, but not both. **(C)** Observed and expected frequencies of the mixed labelling state for all amino acids for supplement concentrations which result in the p closest to 0.5. The depletion of the mixed labelling state varies between amino acids and tends to be generally lower for the mixed state.

calculated as a binomial distribution with two independent trials:

$$b_k = \binom{2}{k} p^k (1-p)^{2-k}$$

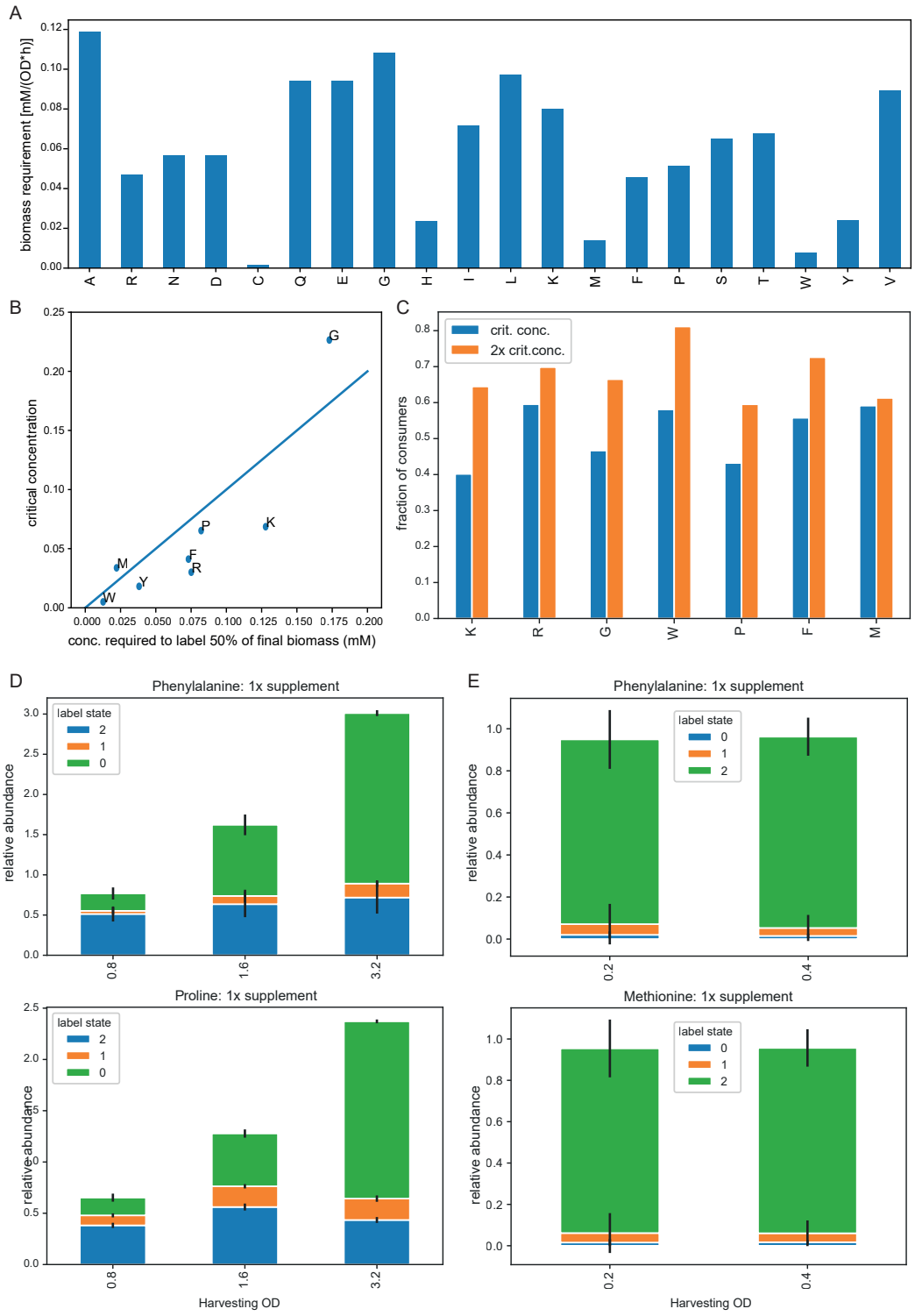
In the case of being near the critical concentration, where $p \approx 0.5$, the mixed labelling states (one of the two residues labelled) would be expected to make up half of the abundance, ie to be together twice as abundant as each of the pure labelling states (Fig. 3.6A, see also Fig. 3.1A). A depletion of this mixed labelling state is clear evidence that cells were relying solely on uptake or synthesis at the time when the protein was made (Fig. 3.6B). Across the 15 amino acids, many show a strong depletion of the mixed labelling state (Fig. 3.6C) and notably, this effect is stronger for the switching amino acids.

3.2.4 Temporal dynamics of uptake patters

The data presented above investigates the total label incorporation into the proteome during growth and implicitly assumed a steady state producer/consumer ratio. However, if the supplement concentration changes significantly during the incubation time, the observed producer/consumer populations could be an artefact of the depletion of the supplement. Rather than coexisting in parallel, producer/consumer sub-populations could exist temporally separated, ie the consumer peptides could have been synthesised early while the consumer peptides were synthesised late. The measurements presented above were done in batch culture and cells were harvested at an OD of 1.6. This is far below the final optical density in this media (which is above 4), so the major nutritional components (glucose, ammonium, phosphate, sulphate etc) are not exhausted. However, the observed critical concentrations of switcher amino acids are very low, typically in the low micromolar range. It is therefore possible and likely that the extracellular concentration changed during growth. In order to estimate the expected concentration change, I calculated the concentration of each amino acid per OD of biomass (C_M) as

$$C_M = f_M * d * p * \frac{1}{w}$$

where f_M is the molar fraction of each amino acid M in the proteome (Lange and Heijnen, 2001), $d \approx 0.667$ is the dry weight per OD unit (approximately 0.67 (mg/mL)/OD),p is the proportion of proteins in the dry biomass (40%w/w) (Yamada and Sgarbieri, 2005) and w is the average amino acid molar mass across the proteome (average of molecular


A
biomass requirement [mM/(OD*h)]
A R N D C Q E G H I L K M F P S T W Y V
B
critical concentration
conc. required to label 50% of final biomass (mM)
C
crit. conc.
2x crit.conc.
fraction of consumers
K R G W P F M
D
Phenylalanine: 1x supplement
label state
relative abundance
Proline: 1x supplement
Harvesting OD
E
Phenylalanine: 1x supplement
Methionine: 1x supplement
Harvesting OD

weights weighted by molar fraction in the proteome). The estimated required amount (ignoring intracellular pools, stores and metabolic pathways consuming amino acids other than translation) varies substantially between amino acids (Fig. 3.7A).

From C_M, it is possible to calculate the supplement concentration required to achieve 50% labelling of the biomass ($C_M * 1.6/2$, with 1.6 being the OD at the time of sampling). This prediction correlates strikingly well with the critical concentrations obtained for switcher amino acids (Fig. 3.7B). This observation is compatible with a model where all cells are consumers until the supplement runs out and then switch to consumers. At the critical concentration, this switch would occur exactly one doubling before sampling, resulting in half the biomass being labelled. Assuming this is the case, one would expect no producer cells if the supplement concentration was doubled, because there would be enough of it to label the entire biomass. Indeed there is usually a substantial increase in the fraction of consumer peptides when the supplement concentration is doubled, although consumer peptides are always still present (Fig. 3.7C).

In order to investigate the temporal dynamics of supplement consumption, I grew bigger cultures supplemented with the critical concentrations of phenylalanine and proline. Samples were taken at OD 0.8, 1.6 and 3.2, ie one doubling before and after the usual sampling time. Relative peptide abundances were quantified as previously. At 1x phenylala-

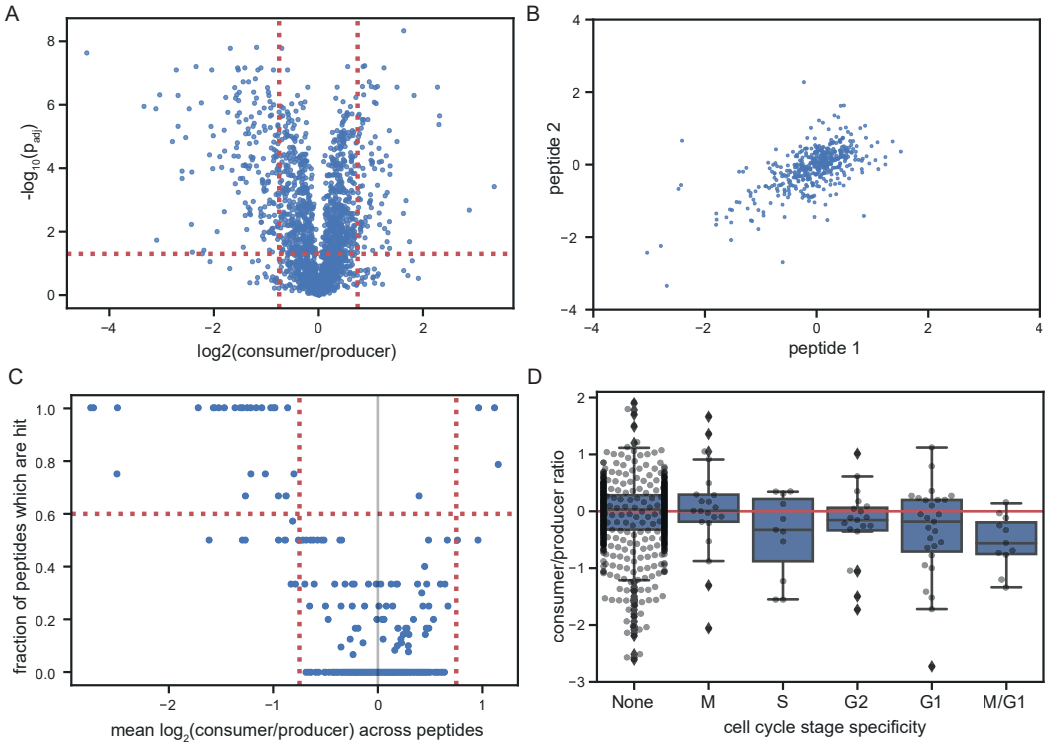
Fig. 3.7 ***(preceding page)*****: Uptake patterns change as the supplement is consumed.** **(A)** Amino acid concentrations per OD of biomass, estimated from literature values (see main text). **(B)** Critical concentrations of switching amino acids (Fig. 3.3B) versus the concentration required to label half the biomass. Critical concentrations are in the range of what would be expected to be consumed to reach the observed level of labelling, ie supplement concentrations can be expected to change very significantly during growth. Both parameters correlate well which raises the possibility of critical concentrations being driven by exhaustion of the media. The blue line shows y=x. **(C)** In a model were all cells are consumers until the supplement is exhausted and then switch to being consumers, doubling the supplement concentration would be expected to abolish producer populations. Shown are the fractions of consumers for 1x and 2x of the supplement concentration closest to the estimated critical concentration. Doubling the supplement concentration generally increases the consumer fraction but does not abolish it. **(D)** Timecourse experiment for F and P supplemented at the critical concentration. Bars were scaled to the optical density at the time of harvesting, making abundances of labelling states directly comparable across timepoints. **(D)** Timecourse experiment for F and M supplemented at the critical concentration. Cultures were sampled at very low ODs to minimise depletion of the supplement. The presence of essentially exclusively consumer peptides at these early timepoints indicate a later switch due to supplement exhaustion.

nine supplementation, at timepoint 1, a significant producer population (27%) was already present. However, the frequency of consumer peptides decreased substantially during the next doubling (from 65% to 38%), and again with the next doubling (to 23%). Peptide ratios were then scaled to the OD at the time of harvesting (Fig. 3.7D). This way, relative abundances become directly comparable, assuming protein turnover is negligible compared to growth and *de novo* synthesis. The scaled consumer population only increased very slightly from T1 ro T2 (from 0.51 to 0.63), indicating that the vast majority of consumer peptides usually found at OD 1.6 were not synthesised during the last doubling. Results for proline supplementation were similar in principle. The producer ratio increased from 23% to 34% between timepoint 1 and 2 and the scaled consumer abundance increased from 0.17 to 0.51. These results indicate clearly that producer/consumer ratios were not in steady state in the cultures between the chosen timepoints. To further investigate, I harvested cultures at even lower ODs (0.2 and 0.4) to minimise the relative concentration change of the supplement. Cultures were started with the critical concentration of phenylalanine and methionine. Methionine was chosen for this experiment because its switching curve displayed an unusual plateau (Fig. 3.2). At these low ODs, both cultures essentially exclusively contained consumer peptides (Fig. 3.7E).

These observations are crucial for interpreting switching curves. Unlike initially intended, switching curves do not reflect distinct, co-existing producer and consumer populations. Instead, at least for the switcher amino acids, they reflect primarily the rate at which the supplement is consumed and thereby at what time it runs out. Producer/consumer populations hence exist, but only temporally separated. This also explains the de-enrichment of the mixed labelling state. The observation that supplementation at twice the critical concentration does not abolish producer populations entirely (Fig. 3.7C) and the existence of some mixed labelling states (Fig. 3.6C) is most likely explained by storage effects. As the internal free amino acid pool turns over at a certain rate, there is a gradual switch when the supplement runs out.

3.2.5 Deconvoluting the proteome of sub-populations

Even though no heterogeneity in amino acid uptake could be detected in the biological context investigated, I believe the developed method and concept still represent meaningful advances for the field. Going one step beyond the previous investigations of a handful of peptides, SWATH proteomics could enable quantification of the labelling state of thousands



Source	Term ID	Name	p	Term size	Query size	Intersection size
KEGG	KEGG:00500	Starch and sucrose metabolism	3.28E-05	14	31	10
GO:BP	GO:0005975	carbohydrate metabolic process	0.005974	51	31	15
GO:BP	GO:0006091	generation of precursor metabolites and energy	0.019012	55	31	15

Fig. 3.8: Deconvoluting producer/consumer proteome signatures. Proteome-wide profiles of 8 biological replicates grown with the critical concentration of phenylalanine were measured by DIA-SWATH-MS. **(A)** Volcano plot illustrating differential labelling in peptides containing at least one phenylalanine (1970 peptides in total). 259 peptides were found to be differentially labelled, using a significance threshold of 0.05 after Benjamini-Hochberg correction (1-sample, two-sided t-test) and an effect size cut-off of ± 0.75. **(B)** Labelling ratios of peptide pairs from the same protein correlate albeit not very strongly (Pearson r=0.62). **(C)** Differential labelling analysis at the protein level, considering only the 436 proteins with at least two detected peptides. A protein was considered a hit if the absolute median log2-transformed labelling ratio exceeded 0.5 and if at least 60% of the individual peptides were hits according to the criteria from (A), yielding 31 hits in total. **(D)** Gene enrichment analysis of hits (versus 436 background proteins) performed with gProfiler (Reimand et al., 2007) using default settings. Genes annotated to central carbon and energy metabolism are strongly enriched in the hits. **(E)** Analysis of labelling patterns in proteins expressed in specific stages of the cell cycle. All proteins with at least one peptide containing at least one phenylalanine were used for this analysis. No clear difference between cell cycle stages is apparent, suggesting that producer/consumer status does not depend on cell cycle stage. Protein annotations were retrieved from Spellman et al. (1998). The y-axis was cut to exclude one low outlier in the None category for visualisation purposes.

of peptides in parallel. This would allow for a much deeper, investigation of sub-populations as long as these can be differentially labelled. By looking for differences in labelling patterns between peptides, and ultimately proteins, their abundance in the producer and consumer sub-populations could be determined, thereby deconvoluting the proteome profiles of the two sub-populations using a simple, intrinsically controlled measurement. This is potentially challenging as the complexity of the peptide mass space increases dramatically when ^{13}C labels are introduced, eg a peptide with 100 carbon atoms has possible 2^{100} isotopomers. Analyses were therefore restricted to only fully labelled and unlabelled amino acids and all but the supplemented amino acid were assumed to be fully labelled.

As a pilot experiment, I grew 8 biological replicates with the critical concentration of phenylalanine and sampled at the usual OD of 1.6. Each sample was injected three times, each time analysing a different mass range of precursors (gas phase fractionation) with a very small (2Da) SWATH precursor isolation window. Data were analysed in DIA-NN, which searches for peptide signatures based on a fasta proteome digested *in silico* and the user-defined fixed and variable modifications. This resulted in 1970 peptides with at least one phenylalanine consistently detected across the replicates (in 6 or more out of the 8 samples). For each sample, the labelling ratio of each peptide was determined, median-centred and log2 transformed. Median labelling ratios ranged between 1.009 and 1.148, close to 1, the expected value at the critical concentration. For each peptide it was then tested if the replicates deviated significantly from 0 (one-sample, two sided t-test) and proteome-wide p-values were corrected for multiple testing using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). From 1970 peptides overall, 259 were found to be significantly ($p_{\text{adj}} < 0.05$) and substantially (± 0.75) different from 0, ie more abundant in the producer or consumer (Fig. 3.8A). For 436 proteins, at least two peptides with at least one phenylalanine each were detected. Labelling ratios of peptide pairs from the same protein correlated reasonably well (Pearson r=0.62), indicating a true biological signal in the data despite some analytical inaccuracies (Fig. 3.8B). Due to this inconsistency, I chose a conservative approach for aggregating the peptide data at the protein level. Only proteins for which at least two peptides with more than one phenylalanine each were measured were considered. A protein was determined to be differentially expressed in producers and consumers if at least 60% of the individual peptides were considered a hit according to the criteria outlined above and if the absolute median of the peptide ratios was greater than

0.5. Out of the 436 proteins, 31 were found to be hits, with 3 being more abundant in the producer and 28 more abundant in the consumer (Fig. 3.8C). The three gene groups upregulated in the producer were all associated with metabolism, namely Arg1, Lys9;Lys13 and Ald6;Aldh1. Strikingly, the general amino acid permease Gap1 was found to be downregulated in the producer (log2-ratio=-1.72, two peptides identified, both hits). Three proteins (Pdc1, His5,Aro8) annotated to phenylalanine biosynthesis or degradation (from SGD pathways) were detected but not found to be differentially labelled. Differentially labelled proteins were significantly enriched in functions in central carbon metabolism and specifically the KEGG pathway 'Starch and Sucrose Metabolism' (Fig. 3.8E). These results show that the deconvolution of subpopulation proteome profiles from SWATH data is possible using this ^{13}C labelling approach. While this represents a significant technological advancement, the biological findings need to be interpreted in combination with the considerations laid out in the previous section. On average, producer peptides have been synthesised later in the growth curve, so that these changes could reflect population-level changes in gene expression in response to glucose depletion, rather than direct effects of phenylalanine production/consumption. This is consistent with the observation that several stress and respiration proteins are upregulated in the producers (namely Ssa1, Gre3, Hsp104 and Cor1, Prx1, Atp3, Cox13), which is consistent with cells entering di-auxic shift.

Nevertheless, I believe these results represent an important proof-of-concept and lay the foundation for exciting future work. By being able to link quantitative proteome profiles with metabolic characteristics could potentially lead to an advanced understanding of what is driving and what are the consequences of metabolic differentiation. This represents a significant advantage compared to other methods such as single-cell sequencing (which can detect sub-populations but not directly link them to a metabolic phenotype) or single-cell metabolomics methods (which fail to characterise the broader physiological state of the cell beyond a few metabolic parameters). To illustrate, I systematically checked the producer/consumer ratio in proteins encoded by genes known to be expressed in specific stages of the cell cycle (Spellman et al., 1998). No difference in labelling pattern is observable, suggesting the phenylalanine uptake behaviour does not depend on cell cycle stage (Fig. 3.8D). If a biological system with true co-existing producer/consumer sub-populations could be constructed, investigations like this could be performed for protein markers characteristic of cell age, activation of the cellular stress response or activity of specific metabolic

pathways such as respiration, conclusively linking a metabolic phenotype with the wider physiological state of the cell.

3.3 Discussion

Amino acid exchange interactions in clonal, exponentially growing and well-mixed cultures -the standard growth condition in the lab- have so far proven elusive. This work represents a novel approach, building on concepts previously developed in the Ralser laboratory (Campbell et al., 2018). Using ^{13}C labelling, the peptides of specific sub-populations are marked when they are synthesised, allowing for the quantification and proteome-wide deconvolution of sub-population proteomes. The central result is the in-depth characterisation of the concentration-dependent uptake behaviour of individual amino acids. Here, the key finding is that the supplementation of some, but not all amino acids completely suppresses self-synthesis, even at very low micro-molar concentrations. These switcher amino acids are generally larger molecules, not preferred nitrogen sources and are lowly abundant in the proteome. However, how self-synthesis is repressed mechanistically is at this moment unclear. Preliminary data (not shown) indicates that the regulation of gene expression is very unlikely to be the reason, as the supplementation of most amino acids causes no or only a mild downregulation of its synthesising enzymes. The observed correlation with the harvesting factor (Fig. 3.4) gives a clue to a potential mechanism where the intracellular accumulation of the supplemented amino acid suppresses its synthesis. This would be an attractive, very dynamic mechanism allowing rapid de-repression when the supplement runs out. I believe this represents a promising avenue for future research, highlighting the dynamic nature of metabolism.

No heterogeneous amino acid utilisation could be detected in the chosen biological context. This could mean that BY4741 yeast cells in minimal media generally do not behave heterogeneously in their amino acid uptake behaviour. Instead, through the expression of high affinity transporters such as Gap1, any lowly concentrated amino acids in the media are scavenged and taken up until exhausted. In other words, there is no active switch from synthesis to uptake. It is well possible that other yeast strains and species differ in this respect. Members of the *S. cerevisiae* species display evidence of high variation and recent selection for nitrogen source utilisation (Ibstedt et al., 2015). The reported concentrations and even the classification of amino acids into switchers/non-switchers might therefore differ, even between different laboratory strains. For example, it has been described that the

laboratory strain Σ1278b suppresses the use of non-preferred nitrogen sources more strongly than the S288c-derived BY-strains when grown on ammonium (Godard et al., 2007). With important groundwork completed, it would be very exciting to apply this method to a range of other yeast and bacterial strains to check if they differ in their behaviour and display heterogeneity.

Another possibility is that BY4741 does respond heterogeneously to amino acids after all, but this was not detected using the chosen method. A key weakness of it is that the initial supplement concentration cannot be lower than the biosynthetic requirement of the culture, otherwise the supplement will run out, leading to an artificial temporal switch from uptake to synthesis. Proteomics analysis requires a substantial amount of biomass, so initial supplement concentrations have to be in the range of tens of micro-molar. This means that heterogeneity could still be induced by concentrations below this limit, inaccessible by batch culture techniques. Using lower sampling ODs (which require bigger culture volumes) can ameliorate this problem, but only to an extent. An alternative possible solution would be continuous culturing devices such as turbidostats and chemostats, but these usually require large amounts of media making them unsuitable for ^{13}C labelling experiments. My experiments have conclusively shown that amino acids are taken up even when present at miniscule concentrations. This opens up the possibility that cooperative metabolite exchange could occur at very low extracellular concentrations, requiring new experimental approaches.

Finally, the reductionist approach taken with this project might be unsuitable for recapitulating metabolic heterogeneity in more natural settings. In this setup, amino acids have to be supplemented one at a time, otherwise peptide masses become impossible to interpret. However, it could be that a combination of amino acids and other factors are required to induce producer/consumer sub-populations. In particular, one would expect that high cell densities favour metabolite exchange, where the dilution effect on exported amino acids is less strong. It could therefore be that amino acid export and uptake heterogeneity is triggered by density-dependent factors such as quorum sensing molecules or the presence of other overflow metabolites such as ethanol, acetate or glycerol. While this project could not find evidence for heterogeneous amino acid metabolism, it has established a novel method for its investigation. Furthermore, the negative result obtained with the BY4741 strain will inform future experiments looking for heterogeneity. In particular, future studies should

investigate the possibility of metabolite exchange at very low extracellular levels and screen for other strains and species displaying heterogeneity.

3.4 Materials and Methods

3.4.1 Strains and reagents

The 'wild type' strain used for this project was based on strain BY4741 (a close relative of the strain S228c used for the *Saccharomyces* genome project). BY4741 is haploid, of mating type MATa and had four key metabolic genes (His3, Leu2, Met15, Ura3) disrupted, making it auxotrophic for histidine, leucine, methionine and uracil (Baker Brachmann et al., 1998). The BY4741 strain was used for generating the first and most popular yeast knock-out collection (Giaever et al., 2002) so a wealth of functional genomics data is available from our lab and others. To construct the strain used in this study, the four auxotrophic markers of strain BY4147 were re-replaced by the endogenous gene in their native loci which generated a prototroph also referred to as BY4147 HLUM knock-in (Alam et al., 2016).

Minimal (SM) media was prepared from yeast nitrogen base without amino acids (YNB, Sigma Y0626) at the concentration recommended by the manufacturer (6.7g/L). All media was filter-sterilised, not autoclaved, prior to use and stored in the dark (because of light-sensitive vitamins contained in YNB). Amino acids were obtained from Sigma and stocks were prepared in water and stored at 4°C up to several weeks. Fully labelled ^{13}C glucose was obtained from Sigma (389374), prepared as a 10% (10x) stock and stored at 4°C. Yeast were grown in 96 polypropylene (PP) deep well plates (Eppendorf), covered by a gas-permeable seal (Breathe-Easy sealing membrane, Sigma Z380059) and grown at 30°C and 1050 rpm shaking (Heidolph Titramax 1000).

3.4.2 Proteomics

Proteomics sample preparation. This protocol for sample preparation from exponentially growing yeasts was originally developed by Christoph Messner. I made a few, small changes to adapt it to reduced input biomass. The protocol is based on a standard reduction alkylation of cysteine residues (Müller and Winter, 2017) followed by tryptic digestion and purification of peptides by solid phase extraction (SPE). Proteomic samples were prepared by adding a small amount of glass beads and 200µl lysis buffer (7M urea, 0.1M ammonium bicarbonate) to frozen pellets followed by mechanical lysis for 5 minutes at 1500rpm (SPEX

MiniG 1600). 20µl of a solution of 55mM dithiothreitol (DTT) was added to each well, the plate shaken for 2 minutes (1000rpm, Eppendorf Thermomix), centrifuged for 4 minutes at 3220g to remove foam and then incubated at 30°C for 60 minutes. The plate was then cooled on ice for 5 minutes, 20µl of 120mM 2-iodoacetamide (IAA) was added and mixed by shaking for 2 minutes, centrifuged for 4 minutes at 3220g to remove foam, followed by 30 minutes at room temperature in the dark. 1ml of 0.1M ammonium bicarbonate was added, the plate was mixed for 1.5 minutes and centrifuged for 3 minutes at 3220g. 920µl of lysate were transferred to a fresh plate and 2µg of modified sequencing grade trypsin (Promega, prepared according to manufacturer's instructions) was added, shaken for 5 minutes and incubated over night at 37°C. 96µl of 10% formic acid were added, followed by shaking for 10 seconds at 1200rpm. Peptides were purified by solid phase extraction using a C18 column (BioPure Macro 96, PROTO 300 C18. Part HNS S18V-L, The Nest Group), which was first conditioned with 200µl of Methanol, followed by twice 200µl of buffer A (50% acetonitrile) and finally three times 200µl of buffer B (3% acetonitrile and 0.1% formic acid). 1000µl of digest were loaded (in two steps of 500µl each), washed three times with buffer B and eluted with twice 120µl buffer A followed by 130µl buffer A. The sample was evaporated (30°C, V-AQ setting, Eppendorf Concentrator plus) and reconstituted in 50µl of buffer B. Sample were centrifuged (5 minutes, 3,200g) and transferred to a fresh plate compatible with our LC autosampler. If required, samples were stored at -80°C until further analysis.

Proteomics data acquisition. Samples were analysed on an ABSciex QTOF 6600 (with *Analyst TF* v1.7.1) coupled to a Waters Acquity UPLC operating at 5µl per minute. 2µg of sample were injected and separated on a C18 reverse phase column (Waters Acquity UPLC HSS T3, 1.8µm) using a 20 minute chromatographic gradient. This method differs significantly from more traditional proteomics methods which use nano-flow chromatography and much longer gradients (typically 1-2 hours). Longer gradients result in better peak separation and wider peaks, allowing longer cycle times, but our micro-flow setup clearly outperforms non-flow methods in terms of throughput and stability (Vowinckel et al., 2018). Electrospray ionisation (ESI) was used to charge and transfer peptides into the gas phase. Source parameters were as follows: ion source gas 1: 15; ion source gas 2: 15; curtain gas 20; temperature: 0; ionspray voltage floating: 5500; declustering potential: 80. Peptides were measured in positive mode and high sensitivity mode throughout all experiments.

Targeted methods were designed using *Skyline* v 20.1 (MacLean et al., 2010) and included in every cycle the acquisition of an MS1 spectrum followed by up to 32 product ion spectra (parallel reaction monitoring). The accumulation time for MS1 measurements was 250ms. The accumulation time for MS2 measurements was 40ms, resulting in a total cycle time remained below 3 seconds. This results in typically at least 10 datapoints per peak allowing precise quantification of peak areas. Collision energies and declustering potentials were predicted using the SCIEX setting in *Skyline*. Proteome-wide labelling patterns were analysed using SWATH acquisition. Each sample was injected three times and a different precursor mass range was analysed each time (gas phase fractionation). Each injection covered a 200 m/z window, from 400-600, 600-800 and 800-1000. Each cycle lasted 2.1 seconds and comprised MS1 scan with 50ms accumulation time and 100 SWATH windows with width 2. Fragmentation parameters were predicted automatically using default settings.

Data analysis. Data acquired by parallel reaction monitoring were analysed in *Skyline*. Only confidently identified peptides with typically at least 3 high quality fragments with a suitably high signal-to-noise ratio were used for further analysis. Fragments were additionally selected to allow for the quantification of all precursor labelling states. To illustrate, precursors with two residues of a single supplemented amino acid can have four different labelling states, two pure states and two mixed states. However, the mixed states have the same precursor mass, so cannot be distinguished at the MS1-level. Accordingly, only fragments unique to one of the mixed labelling states can be used for quantification and these are fragments which have one and only one of the supplemented amino acid. Peaks were usually integrated using fixed retention time windows. These were manually curated and then applied to all precursors belonging to the same peptide group across all samples. While this method does not always give perfectly integrated peaks when retention times shift during sample batches, it strictly compares signal intensities across identical time windows, which is crucial when computing the ratio between labelling states. Peptide quantification reports were generated in *Skyline* and analysed further in Python.

DIA data was analysed using *DIA-NN* v1.7.12 (Demichev et al., 2020) with the following settings different to defaults: MS2 resolution:20,000; MS1 resolution:12,000; scan window:8. Variable and fixed modifications of type 'label' were used as appropriate according to the expected labelling state of the analysed peptides.

3.4.3 Metabolomics

Four colonies of the wild type strain were picked and grown in 10-ml SM pre-cultures overnight. Cultures were diluted to OD0.2 and grown for 6 hours in fresh SM media in the presence of single amino acids (0.334mM) or yeast extract (2%). Amino acids were extracted and quantified as previously established by Mülleder et al. (2016). In brief, pellets were washed twice to remove extracellular amino acids from the pellet. 100µl hot ethanol (80% w/w, 80°C) was added and incubated at 80°C for 2 minutes, followed by vigorous mixing and another 2 minutes at 80°C. Soluble extracts were separated by centrifugation (3,200g, 5min). 1µl of each sample was analysed by LC-MS/MS. Compounds were separated by hydrophilic interaction chromatography (HILIC) using a Waters ACQUITY UPLC BEH amide 1.7 mm 2.1 × 100 mm column at 25°C. The buffers used were: Buffer A 50:50 acetonitrile/water, 10 mM ammonium formate, 0.176% formic acid, and Buffer B 95:5:5 acetonitrile/methanol/water, 10 mM ammonium formate and 0.176% formic acid. Flow was maintained at 0.9 ml per minute starting with 0.7 min at 85% buffer B, followed by a 1.85-min ramp leading to 5% B which was kept constant for 0.05 min before returning to the starting conditions. Multiple reaction monitoring (MRM) on a triple quadrupole mass spectrometer (Agilent 6460) was used for quantification. Transition settings were imported from (Mülleder et al., 2017). A standard curve was prepared by serial dilution of pure amino acids and used to convert peak areas to absolute concentrations.

Conclusions

This thesis has covered three projects investigating yeast metabolism and physiology. Each used yeast as a model for eukaryotic metabolism and employed high-throughput or genome-wide approaches, as well as a combination of wet and dry techniques. By working on such a multi-faceted set of projects over the last four years, I have become proficient in liquid chromatography - mass spectrometry, laboratory automatisation, statistical analysis and software development in python, as well as many standard microbiological and biochemical techniques. While this was an immensely stimulating and educational experience in itself, I hope that each of my published findings has advanced the field's understanding and methods. In particular, the key take-home messages that have emerged from this work are:

- *Pyphe* is a new, all-in-one solution for colony fitness screening
- Colony redness analysis by phloxine B staining provides an orthogonal and independent readout to colony size
- Little is gained from recording colony growth curves compared to classical endpoint measurements
- A SNP in *pyk1* affects growth, chronological lifespan and stress resistance in *S. pombe* wild strains
- The *S. pombe* standard lab strain carries the unusual low activity Pyk1 isoform
- A subset of amino acids is taken up at high affinity and completely suppress cell-internal synthesis in *S. cerevisiae*
- Heterogeneous amino acid metabolism could be detected by analysing ^{13}C labelling patterns in peptides

The long-term success of *pyphe* will depend on the formation of a critical mass of users and the continuous support and development of its interface and algorithms. Hopefully, it will contribute towards unifying and standardising data analysis workflows across laboratories, even if researchers end up using *pyphe* only for parts of their analysis pipeline. With this in mind, I believe the modular design of *pyphe* is its biggest strength which makes it preferable to large, complicated GUI-driven solutions. Using *pyphe*, I have investigated optimal phenotyping strategies and come to the conclusion that probably little is gained from timecourse imaging of colonies. This is perhaps somewhat surprising, given that recently published methods on this topic all appear to favour recording colony growth curves (Zackrisson et al., 2016; Banks et al., 2012; Levin-Reisman et al., 2014). But contrary to Zackrisson et al. (2016), my work focuses not solely on the level of technical noise in the readout but on the nature of the information contained in endpoint and timecourse measurements. The surprising finding is that maximum slopes correlate very tightly with endpoints, not something that would be expected from analogous growth in liquid cultures. There, the substrate conversion efficiency determines the final biomass. On agar however, colonies have a limited time window for growth that is determined by the ensemble of colonies exhausting the growth media and colony size is thereby determined primarily by the growth rate during that time.

Pyphe was then employed in a large-scale functional genomics screen of lincRNA function, with mixed success. We were able to uncover phenotypes for approximately half of the lincRNAs tested, which is a significant achievement given the sparseness and subtlety of their phenotypes. This will provide a valuable resource for further, focused investigation of individual lincRNAs, especially once this dataset is integrated into PomBase, *S. pombe's* model organism databank (Wood et al., 2012). However, the data was in the end not sufficient to perform conclusive clustering analyses. This was partly due to the fact that only a small subset of protein-coding genes was contained in the dataset, not enough to observe fine-grained, statistically significant enrichments and clustering of related functional categories in the obtained clusters. A new, unpublished dataset by María Rodrígues-López is going to address this shortcoming by measuring the entire non-essential deletion collection in a large number of conditions.

I believe my work on amino acid uptake and heterogeneity (Chapter 3) represents a potentially very useful and novel approach to a key knowledge gap in the field. While we are

beginning to understand the species composition, pan-genomes and pan-transcriptomes of microbial communities through ever deeper sequencing, the exact interplay and function of individual community members is often not understood. One popular approach has been to reconstruct communal genome-scale metabolic networks from sequencing data, which are then amenable to constraint-based modelling (Sen and Orešič, 2019). However, these models describe what individual members can do (based on genomically-encoded enzymes), but not what they actually do in practice (which is almost certainly subset of the former). My work on amino acid uptake patterns shows very clearly that lowly concentrated extracellular amino acids are not only taken up in ammonium-replete conditions, but that a subset of them completely suppresses cell-internal biosynthesis in these prototrophic cells. Here, a key objective of future investigations will be to elucidate the mechanism by which cells switch off endogenous synthesis.

Analogously, in natural settings, we would expect yeast (and potentially other microbes) to suppress the synthesis of certain compounds available to them in their surrounding. This has profound implications beyond the immediate metabolic response. We have previously shown that the inactivation of specific biosynthetic pathways (by plasmid loss) can have strong effects on cellular stress resistance in *S. cerevisiae* (Campbell et al., 2016). More recent, our unpublished results show a wide-spread modulation of antifungal drug tolerance and resistance by amino acid supplementation. Understanding the effect that specific supplements and the metabolic environment of the niche in general have on intracellular metabolism will therefore contribute to the development of new anti-microbial therapeutics. Approaches like mine, combined with a broad investigation of intracellular changes, will help understand how these potential new treatments exert their effect. This is particularly important in the light of the worrying spread of antimicrobial resistance and multi-drug resistant strains (Zaman et al., 2017).

Based on these observations, it is reasonable to expect that if infectious microbes were metabolically heterogeneous, they would show an associated heterogeneity in drug resistance and/or tolerance. Such heterogeneous resistance effects are well described and manifest themselves in drug-resistance subpopulations which evade treatment (Rosenberg et al., 2018). If a causative relationship with metabolic heterogeneity could be established, this would represent a significant achievement that could result in treatments that suppress metabolic heterogeneity, and in turn render the entire population sensitive to the drug. To

this end, my method could be used to probe for heterogeneous amino acid metabolism in other microbial species, especially clinically relevant ones such as *Candida albicans*, and to eventually link this heterogeneity to other phenotypes.

References

Addinall, S. G., Holstein, E.-M., Lawless, C., Yu, M., Chapman, K., Banks, A. P., Ngo, H.-P., Maringele, L., Taschuk, M., Young, A. et al. (2011). Quantitative fitness analysis shows that NMD proteins and many other protein complexes suppress or enhance distinct telomere cap defects. *PLoS Genet.* 7: e1001362.

Alam, M. T., Olin-Sandoval, V., Stincone, A., Keller, M. A., Zelezniak, A., Luisi, B. F. and Ralser, M. (2017). The self-inhibitory nature of metabolic networks and its alleviation through compartmentalization. *Nature communications* 8: 1–13.

Alam, M. T., Zelezniak, A., Mülleder, M., Shliaha, P., Schwarz, R., Capuano, F., Vowinckel, J., Radmaneshfar, E., Krüger, A., Calvani, E. et al. (2016). The metabolic background is a global player in saccharomyces gene expression epistasis. *Nature microbiology* 1: 15030.

Allert, S., Ernest, I., Poliszczak, A., Opperdoes, F. R. and Michels, P. A. (1991). Molecular cloning and analysis of two tandemly linked genes for pyruvate kinase of trypanosoma brucei. *Eur. J. Biochem.* 200: 19–27.

Anastasiou, D., Poulogiannis, G., Asara, J. M., Boxer, M. B., Jiang, J.-k., Shen, M., Bellinger, G., Sasaki, A. T., Locasale, J. W., Auld, D. S. et al. (2011). Inhibition of pyruvate kinase m2 by reactive oxygen species contributes to cellular antioxidant responses. *Science* 334: 1278–1283.

Andersen, K. B. and von Meyenburg, K. (1980). Are growth rates of escherichia coli in batch cultures limited by respiration? *J. Bacteriol.* 144: 114–123.

Ata, Ö., Rebnegger, C., Tatto, N. E., Valli, M., Mairinger, T., Hann, S., Steiger, M. G., Çalık, P. and Mattanovich, D. (2018). A single gal4-like transcription factor activates the crabtree effect in komagataella phaffii. *Nat. Commun.* 9: 4911.

Atkinson, D. E. and Walton, G. M. (1967). Adenosine triphosphate conservation in metabolic regulation. rat liver citrate cleavage enzyme. *J. Biol. Chem.* 242: 3239–3241.

Atkinson, S. R., Marguerat, S. and Bähler, J. (2012). Exploring long non-coding rnas through sequencing. In *Seminars in cell & developmental biology*, volume 23, pp. 200–205. Elsevier.

Atkinson, S. R., Marguerat, S., Bitton, D. A., Rodríguez-López, M., Rallis, C., Lemay, J.-F., Cotobal, C., Malecki, M., Smialowski, P., Mata, J. et al. (2018). Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. *RNA* 24: 1195–1213.

Baker Brachmann, C., Davies, A., Cost, G. J., Caputo, E., Li, J., Hieter, P. and Boeke, J. D. (1998). Designer deletion strains derived from saccharomyces cerevisiae s288c: a useful set of strains and plasmids for pcr-mediated gene disruption and other applications. *Yeast* 14: 115–132.

Balarezo-Cisneros, L. N., Parker, S., Fraczek, M. G., Timouma, S., Wang, P., O'Keefe, R. T., Millar, C. B. and Delneri, D. (2020). Functional and transcriptional profiling of non-coding rnas in yeast reveal context-dependent phenotypes and widespread in trans effects on the protein regulatory network. *bioRxiv* .

Banks, A. P., Lawless, C. and Lydall, D. A. (2012). A quantitative fitness analysis workflow. *J. Vis. Exp.* .

Barnett, J. A. (2003). A history of research on yeasts 5: the fermentation pathway. *Yeast* 20: 509–543.

Baryshnikova, A., Costanzo, M., Dixon, S., Vizeacoumar, F. J., Myers, C. L., Andrews, B. and Boone, C. (2010). Synthetic genetic array (SGA) analysis in saccharomyces cerevisiae and schizosaccharomyces pombe. *Methods Enzymol.* 470: 145–179.

Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J. R. and Hwa, T. (2015). Overflow metabolism in escherichia coli results from efficient proteome allocation. *Nature* 528: 99–104.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57: 289–300.

Bianchi, F., van't Klooster, J. S., Ruiz, S. J. and Poolman, B. (2019). Regulation of amino acid transport in saccharomyces cerevisiae. *Microbiology and Molecular Biology Reviews* 83: e00024–19.

Bischof, L., Převorovský, M., Rallis, C., Jeffares, D. C., Arzhaeva, Y. and Bähler, J. (2016). Spotsizer: High-throughput quantitative analysis of microbial growth. *Biotechniques* 61: 191–201.

Björkeroth, J., Campbell, K., Malina, C., Yu, R., Di Bartolomeo, F. and Nielsen, J. (2020). Proteome reallocation from amino acid biosynthesis to ribosomes enables yeast to grow faster in rich media. *Proceedings of the National Academy of Sciences* 117: 21804–21812.

Blacker, T. S. and Duchen, M. R. (2016). Investigating mitochondrial redox state using NADH and NADPH autofluorescence. *Free Radic. Biol. Med.* 100: 53–65.

Blomberg, A. (2011). Measuring growth rate in high-throughput growth phenotyping. *Curr. Opin. Biotechnol.* 22: 94–102.

Bluemlein, K., Grüning, N.-M., Feichtinger, R. G., Lehrach, H., Kofler, B. and Ralser, M. (2011). No evidence for a shift in pyruvate kinase PKM1 to PKM2 expression during tumorigenesis. *Oncotarget* 2: 393–400.

Bradley, P. H., Gibney, P. A., Botstein, D., Troyanskaya, O. G. and Rabinowitz, J. D. (2019). Minor isozymes tailor yeast metabolism to carbon availability. *mSystems* 4.

Brochado, A. R., Telzerow, A., Bobonis, J., Banzhaf, M., Mateus, A., Selkrig, J., Huth, E., Bassler, S., Zamarreño Beas, J., Zietek, M. et al. (2018). Species-specific activity of antibacterial drug combinations. *Nature* 559: 259–263.

Brochado, A. R. and Typas, A. (2013). High-throughput approaches to understanding gene function and mapping network architecture in bacteria. *Current Opinion in Microbiology* 16: 199–206.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.

Campbell, K., Herrera-Dominguez, L., Correia-Melo, C., Zelezniak, A. and Ralser, M. (2018). Biochemical principles enabling metabolic cooperativity and phenotypic heterogeneity at the single cell level. *Current Opinion in Systems Biology* 8: 97–108.

Campbell, K., Vowinckel, J., Mülleder, M., Malmsheimer, S., Lawrence, N., Calvani, E., Miller-Fleming, L., Alam, M. T., Christen, S., Keller, M. A. et al. (2015). Self-establishing communities enable cooperative metabolite exchange in a eukaryote. *Elife* 4: e09943.

Campbell, K., Vowinckel, J. and Ralser, M. (2016). Cell-to-cell heterogeneity emerges as consequence of metabolic cooperation in a synthetic yeast community. *Biotechnology journal* 11: 1169–1178.

Carmel-Harel, O. and Storz, G. (2000). Roles of the glutathione- and thioredoxin-dependent reduction systems in the escherichia coli and saccharomyces cerevisiae responses to oxidative stress. *Annu. Rev. Microbiol.* 54: 439–461.

Chen, D., Toone, W. M., Mata, J., Lyne, R., Burns, G., Kivinen, K., Brazma, A., Jones, N. and Bähler, J. (2003). Global transcriptional responses of fission yeast to environmental stress. *Mol. Biol. Cell* 14: 214–229.

Chiron, S., Gaisne, M., Guillou, E., Belenguer, P., Desmond Clark-Walker, G. and Bonnefoy, N. (2007). Studying mitochondria in an attractive model: Schizosaccharomyces pombe. *Methods in Molecular Biology* pp. 91–105.

Chong, Y. T., Koh, J. L., Friesen, H., Duffy, S. K., Cox, M. J., Moses, A., Moffat, J., Boone, C. and Andrews, B. J. (2015). Yeast proteome dynamics from single cell imaging and automated analysis. *Cell* 161: 1413–1424.

Christen, S. and Sauer, U. (2011). Intracellular characterization of aerobic glucose metabolism in seven yeast species by 13C flux analysis and metabolomics. *FEMS Yeast Res.* 11: 263–272.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X. and Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 6: 80–92.

Conrad, M., Schothorst, J., Kankipati, H. N., Van Zeebroeck, G., Rubio-Texeira, M. and Thevelein, J. M. (2014). Nutrient sensing and signaling in the yeast saccharomyces cerevisiae. *FEMS microbiology reviews* 38: 254–299.

Cook, D. E. and Andersen, E. C. (2017). VCF-kit: assorted utilities for the variant call format. *Bioinformatics* 33: 1581–1582.

Costanzo, M., VanderSluis, B., Koch, E. N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S. D. et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353.

Crabtree, H. G. (1929). Observations on the carbohydrate metabolism of tumours. *Biochem. J* 23: 536–545.

Crépin, L., Nidelet, T., Sanchez, I., Dequin, S. and Camarasa, C. (2012). Sequential use of nitrogen compounds by saccharomyces cerevisiae during wine fermentation: a model based on kinetic and regulation characteristics of nitrogen permeases. *Applied and environmental microbiology* 78: 8102–8111.

Cubillos, F. A., Parts, L., Salinas, F., Bergström, A., Scovacricchi, E., Zia, A., Illingworth, C. J., Mustonen, V., Ibstedt, S., Warringer, J. et al. (2013). High-resolution mapping of complex traits with a four-parent advanced intercross yeast population. *Genetics* 195: 1141–1155.

David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C. J., Bofkin, L., Jones, T., Davis, R. W. and Steinmetz, L. M. (2006). A high-resolution map of transcription in the yeast genome. *Proceedings of the National Academy of Sciences* 103: 5320–5325.

De Filippi, L., Fournier, M., Cameroni, E., Linder, P., De Virgilio, C., Foti, M. and Deloche, O. (2007). Membrane stress is coupled to a rapid translational control of gene expression in chlorpromazine-treated cells. *Curr. Genet.* 52: 171–185.

De la Fuente, I. M., Cortés, J. M., Valero, E., Desroches, M., Rodrigues, S., Malaina, I. and Martínez, L. (2014). On the dynamics of the adenylate energy system: homeorhesis vs homeostasis. *PLoS One* 9: e108676.

de Lorenzo, V., Sekowska, A. and Danchin, A. (2015). Chemical reactivity drives spatiotemporal organisation of bacterial metabolism. *FEMS microbiology reviews* 39: 96–119.

Demichev, V., Messner, C. B., Vernardis, S. I., Lilley, K. S. and Ralser, M. (2020). Diann: neural networks and interference correction enable deep proteome coverage in high throughput. *Nature methods* 17: 41–44.

Drakulic, T., Temple, M. D., Guido, R., Jarolim, S., Breitenbach, M., Attfield, P. V. and Dawes, I. W. (2005). Involvement of oxidative stress response genes in redox homeostasis, the level of reactive oxygen species, and ageing in saccharomyces cerevisiae. *FEMS Yeast Res.* 5: 1215–1228.

Druseikis, M., Ben-Ari, J. and Covo, S. (2019). The goldilocks effect of respiration on canavanine tolerance in saccharomyces cerevisiae. *Current genetics* 65: 1199–1215.

Efeyan, A., Comb, W. C. and Sabatini, D. M. (2015). Nutrient-sensing mechanisms and pathways. *Nature* 517: 302–310.

Ephrussi, B. and Hottinguer, H. and Tavlitzki, J. (1949). Action de l'acriflavine sur les levures. i. la mutation 'petite' colonie. *Ann. Inst. Pasteur* 76: 351–367.

Ferenci, T. (2016). Trade-off mechanisms shaping the diversity of bacteria. *Trends Microbiol.* 24: 209–223.

Fröhlich, F., Christiano, R. and Walther, T. C. (2013). Native silac: metabolic labeling of proteins in prototroph microorganisms based on lysine synthesis regulation. *Molecular & Cellular Proteomics* 12: 1995–2005.

Garrison, E. and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv* .

Gehrig, S., Macpherson, J. A., Driscoll, P. C., Symon, A., Martin, S. R., MacRae, J. I., Kleinjung, J., Fraternali, F. and Anastasiou, D. (2017). An engineered photoswitchable mammalian pyruvate kinase. *FEBS J.* 284: 2955–2980.

Georgeson, P., Syme, A., Sloggett, C., Chung, J., Dashnow, H., Milton, M., Lonsdale, A., Powell, D., Seemann, T. and Pope, B. (2019). Bionitio: demonstrating and facilitating best practices for bioinformatics command-line software. *GigaScience* 8: giz109.

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B. et al. (2002). Functional profiling of the saccharomyces cerevisiae genome. *nature* 418: 387–391.

Giaever, G. and Nislow, C. (2014). The yeast deletion collection: a decade of functional genomics. *Genetics* 197: 451–465.

Giani, A. M., Gallo, G. R., Gianfranceschi, L. and Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal* 18: 9–19.

Godard, P., Urrestarazu, A., Vissers, S., Kontos, K., Bontempi, G., van Helden, J. and André, B. (2007). Effect of 21 different nitrogen sources on global gene expression in the yeast saccharomyces cerevisiae. *Molecular and cellular biology* 27: 3065–3086.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M. et al. (1996). Life with 6000 genes. *Science* 274: 546–567.

Gouy, M., Guindon, S. and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27: 221–224.

Grech, L., Jeffares, D. C., Sadée, C. Y., Rodríguez-López, M., Bitton, D. A., Hoti, M., Biagosch, C., Aravani, D., Speekenbrink, M., Illingworth, C. J. R. et al. (2019). Fitness landscape of the fission yeast genome. *Mol. Biol. Evol.* 36: 1612–1623.

Grüning, N.-M., Rinnerthaler, M., Bluemlein, K., Mülleder, M., Wamelink, M. M. C., Lehrach, H., Jakobs, C., Breitenbach, M. and Ralser, M. (2011). Pyruvate kinase triggers a metabolic feedback loop that controls redox metabolism in respiring cells. *Cell Metab.* 14: 415–427.

Haas, R., Zelezniak, A., Iacovacci, J., Kamrad, S., Townsend, S. and Ralser, M. (2017). Designing and interpreting 'multi-omic'experiments that may change our understanding of biology. *Current Opinion in Systems Biology* 6: 37–45.

Hackett, S. R., Zanotelli, V. R., Xu, W., Goya, J., Park, J. O., Perlman, D. H., Gibney, P. A., Botstein, D., Storey, J. D. and Rabinowitz, J. D. (2016). Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* 354: aaf2786.

Haffter, P. and Fox, T. D. (1992). Nuclear mutations in the petite-negative yeast schizosaccharomyces pombe allow growth of cells lacking mitochondrial dna. *Genetics* 131: 255–260.

Hazelwood, L. A., Daran, J.-M., Van Maris, A. J., Pronk, J. T. and Dickinson, J. R. (2008). The ehrlich pathway for fusel alcohol production: a century of research on saccharomyces cerevisiae metabolism. *Applied and environmental microbiology* 74: 2259–2266.

Heslot, H., Goffeau, A. and Louis, C. (1970). Respiratory metabolism of a 'petite negative' yeast schizosaccharomyces pombe 972h-. *J. Bacteriol.* 104: 473–481.

Hillenmeyer, M. E., Fung, E., Wildenhain, J., Pierce, S. E., Hoon, S., Lee, W., Proctor, M., St. Onge, R. P., Tyers, M., Koller, D. et al. (2008). The chemical genomic portrait of yeast: Uncovering a phenotype for all genes. *Science* 320: 362–365.

Hu, W., Jiang, Z.-D., Suo, F., Zheng, J.-X., He, W.-Z. and Du, L.-L. (2017). A large gene family in fission yeast encodes spore killers that subvert mendel's law. *Elife* 6.

Huberts, D. H. E. W., Niebel, B. and Heinemann, M. (2012). A flux-sensing mechanism could regulate the switch between respiration and fermentation. *FEMS Yeast Res.* 12: 118–128.

Ibstedt, S., Stenberg, S., Bagés, S., Gjuvsland, A. B., Salinas, F., Kourtchenko, O., Samy, J. K., Blomberg, A., Omholt, S. W. and Liti, G. (2015). Concerted evolution of life stage performances signals recent selection on yeast nitrogen use. *Molecular Biology and Evolution* 32: 153–161.

Israelsen, W. J. and Vander Heiden, M. G. (2015). Pyruvate kinase: Function, regulation and role in cancer. *Semin. Cell Dev. Biol.* 43: 43–51.

Ito, A., Sato, Y., Kudo, S., Sato, S., Nakajima, H. and Toba, T. (2003). The screening of hydrogen peroxide-producing lactic acid bacteria and their application to inactivating psychrotrophic food-borne pathogens. *Curr. Microbiol.* 47: 231–236.

Jariani, A., Vermeersch, L., Cerulus, B., Perez-Samper, G., Voordeckers, K., Van Brussel, T., Thienpont, B., Lambrechts, D. and Verstrepen, K. J. (2020). A new protocol for single-cell rna-seq reveals stochastic gene expression during lag phase in budding yeast. *eLife* 9: e55320.

Jeffares, D. C. (2018). The natural diversity and ecology of fission yeast. *Yeast* 35: 253–260.

Jeffares, D. C., Rallis, C., Rieux, A., Speed, D., Převorovský, M., Mourier, T., Marsellach, F. X., Iqbal, Z., Lau, W., Cheng, T. M. K. et al. (2015). The genomic and phenotypic diversity of schizosaccharomyces pombe. *Nat. Genet.* 47: 235–241.

Jiménez, R. C., Kuzak, M., Alhamdoosh, M., Barker, M., Batut, B., Borg, M., Capella-Gutierrez, S., Hong, N. C., Cook, M., Corpas, M. et al. (2017). Four simple recommendations to encourage best practices in research software. *F1000Research* 6.

Kahm, M., Hasenbrink, G., Lichtenberg-Frate, H., Ludwig, J. and Kschischo, M. (2010). Grofit: Fitting biological growth curves. *Nature Precedings* .

Kamrad, S., Grossbach, J., Rodríguez-López, M., Mülleder, M., Townsend, S., Cappelletti, V., Stojanovski, G., Correia-Melo, C., Picotti, P., Beyer, A. et al. (2020a). Pyruvate kinase variant of fission yeast tunes carbon metabolism, cell regulation, growth and stress resistance. *Molecular Systems Biology* 16: e9270.

Kamrad, S., Rodriguez-Lopez, M., Cotobal, C., Correia-Melo, C., Ralser, M. and Bähler, J. (2020b). *Pyphe*, a python toolbox for assessing microbial growth and cell viability in high-throughput colony screens. *eLife* 9: e55160.

Kayne, F. J. (1971). Thallium (i) activation of pyruvate kinase. *Arch. Biochem. Biophys.* 143: 232–239.

Keren, L., Van Dijk, D., Weingarten-Gabbay, S., Davidi, D., Jona, G., Weinberger, A., Milo, R. and Segal, E. (2015). Noise in gene expression is coupled to growth rate. *Genome Research* 25: 1893–1902.

Kim, H., Esser, L., Bilayet Hossain, M., Xia, D., Yu, C.-A., Rizo, J., van der Helm, D. and Deisenhofer, J. (1999). Structure of antimycin a1, a specific electron transfer inhibitor of Ubiquinol Cytochrome c Oxidoreductase. *J. Am. Chem. Soc.* 121: 4902–4903.

Kritikos, G., Banzhaf, M., Herrera-Dominguez, L., Koumoutsi, A., Wartel, M., Zietek, M. and Typas, A. (2017). A tool named iris for versatile high-throughput phenotyping in microorganisms. *Nat Microbiol* 2: 17014.

Kumar, R., Ghosh, M., Kumar, S. and Prasad, M. (2020). Single cell metabolomics: A future tool to unmask cellular heterogeneity and virus-host interaction in context of emerging viral diseases. *Frontiers in Microbiology* 11: 1152.

Kuzmin, E., VanderSluis, B., Wang, W., Tan, G., Deshpande, R., Chen, Y., Usaj, M., Balint, A., Mattiazzi Usaj, M., van Leeuwen, J. et al. (2018). Systematic analysis of complex genetic interactions. *Science* 360.

Lange, H. and Heijnen, J. (2001). Statistical reconciliation of the elemental and molecular biomass composition of saccharomyces cerevisiae. *Biotechnology and bioengineering* 75: 334–344.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R. et al. (2007). Clustal W and clustal X version 2.0. *Bioinformatics* 23: 2947–2948.

Lawless, C., Wilkinson, D. J., Young, A., Addinall, S. G. and Lydall, D. A. (2010). Colonyzer: automated quantification of micro-organism growth characteristics on solid agar. *BMC Bioinformatics* 11: 287.

Leupold, S., Hubmann, G., Litsios, A., Meinema, A. C., Takhaveev, V., Papagiannakis, A., Niebel, B., Janssens, G., Siegel, D. and Heinemann, M. (2019). Saccharomyces cerevisiae goes through distinct metabolic phases during its replicative lifespan. *Elife* 8: e41046.

Levin-Reisman, I., Fridman, O. and Balaban, N. Q. (2014). ScanLag: high-throughput quantification of colony growth and lag time. *J. Vis. Exp.* .

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009). The sequence Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Lie, S., Banks, P., Lawless, C., Lydall, D. and Petersen, J. (2018). The contribution of non-essential schizosaccharomyces pombe genes to fitness in response to altered nutrient supply and target of rapamycin activity. *Open Biology* 8: 180015.

Lin, Y., Li, Y., Zhu, N., Han, Y., Jiang, W., Wang, Y., Si, S. and Jiang, J. (2014). The antituberculosis antibiotic capreomycin inhibits protein synthesis by disrupting interaction between ribosomal proteins L12 and L10. *Antimicrob. Agents Chemother.* 58: 2038–2044.

Liti, G. and Louis, E. J. (2012). Advances in quantitative trait analysis in yeast. *PLOS Genetics* 8: 1–7.

Litsios, A., Ortega, Á. D., Wit, E. C. and Heinemann, M. (2018). Metabolic-flux dependent regulation of microbial physiology. *Curr. Opin. Microbiol.* 42: 71–78.

Ljungdahl, P. (2009). Amino-acid-induced signalling via the SPS-sensing pathway in yeast. *Biochemical Society Transactions* 37: 242–247.

Ljungdahl, P. O. and Daignan-Fornier, B. (2012). Regulation of amino acid, nucleotide, and phosphate metabolism in saccharomyces cerevisiae. *Genetics* 190: 885–929.

Lu, W., Wang, L., Chen, L., Hui, S. and Rabinowitz, J. D. (2018). Extraction and quantitation of nicotinamide adenine dinucleotide redox cofactors. *Antioxid. Redox Signal.* 28: 167–179.

MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C. and MacCoss, M. J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26: 966–968.

Malecki, M. and Bähler, J. (2016). Identifying genes required for respiratory growth of fission yeast. *Wellcome Open Res* 1: 12.

Malecki, M., Bitton, D. A., Rodríguez-López, M., Rallis, C., Calavia, N. G., Smith, G. C. and Bähler, J. (2016). Functional and regulatory profiling of energy metabolism in fission yeast. *Genome Biol.* 17: 240.

Malecki, M., Kamrad, S., Ralser, M. and Bähler, J. (2020). Mitochondrial respiration is required to provide amino acids during fermentative proliferation of fission yeast. *EMBO reports* p. e50845.

Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R. and Bähler, J. (2012). Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* 151: 671–683.

Matynia, A., Mueller, U., Ong, N., Demeter, J., Granger, A. L., Hinata, K. and Sazer, S. (1998). Isolation and characterization of fission yeast sns mutants defective at the mitosis-to-interphase transition. *Genetics* 148: 1799.

Messner, C. B., Demichev, V., Wendisch, D., Michalick, L., White, M., Freiwald, A., Textoris-Taube, K., Vernardis, S. I., Egger, A.-S., Kreidl, M. et al. (2020). Ultra-high-throughput clinical proteomics reveals classifiers of covid-19 infection. *Cell systems* 11: 11–24.

Meunier, J.-R. and Choder, M. (1999). Saccharomyces cerevisiae colony growth and ageing: biphasic growth accompanied by changes in gene expression. *Yeast* 15: 1159–1169.

Molenaar, D., van Berlo, R., de Ridder, D. and Teusink, B. (2009). Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol. Syst. Biol.* 5: 323.

Monteiro, F., Hubmann, G., Takhaveev, V., Vedelaar, S. R., Norder, J., Hekelaar, J., Saldida, J., Litsios, A., Wijma, H. J., Schmidt, A. et al. (2019). Measuring glycolytic flux in single yeast cells with an orthogonal synthetic biosensor. *Molecular systems biology* 15: e9071.

Mori, M., Marinari, E. and De Martino, A. (2019). A yield-cost tradeoff governs escherichia coli's decision between fermentation and respiration in carbon-limited growth. *npj Systems Biology and Applications* 5.

Morris, J. J., Lenski, R. E. and Zinser, E. R. (2012). The black queen hypothesis: evolution of dependencies through adaptive gene loss. *MBio* 3.

Mülleder, M., Bluemlein, K. and Ralser, M. (2017). A high-throughput method for the quantitative determination of free amino acids in saccharomyces cerevisiae by hydrophilic interaction chromatography–tandem mass spectrometry. *Cold Spring Harbor Protocols* 2017: pdb–prot089094.

Mülleder, M., Calvani, E., Alam, M. T., Wang, R. K., Eckerstorfer, F., Zelezniak, A. and Ralser, M. (2016). Functional metabolomics describes the yeast biosynthetic regulome. *Cell* 167: 553–565.e12.

Müller, T. and Winter, D. (2017). Systematic evaluation of protein reduction and alkylation reveals massive unspecific side effects by iodine-containing reagents. *Molecular & Cellular Proteomics* 16: 1173–1187.

Muñoz, M. E. and Ponce, E. (2003). Pyruvate kinase: current status of regulatory and functional properties. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 135: 197–218.

Nadal-Ribelles, M., Islam, S., Wei, W., Latorre, P., Nguyen, M., de Nadal, E., Posas, F. and Steinmetz, L. M. (2019). Sensitive high-throughput single-cell rna-seq reveals within-clonal transcript correlations in yeast populations. *Nature microbiology* 4: 683–692.

Nairn, J., Duncan, D., Gray, L. M., Urquhart, G., Binnie, M., Byron, O., Fothergill-Gilmore, L. A. and Price, N. C. (1998). Purification and characterization of pyruvate kinase from schizosaccharomyces pombe: evidence for an unusual quaternary structure. *Protein Expr. Purif.* 14: 247–253.

Nairn, J., Smith, S., Allison, P. J., Rigden, D., Fothergill-Gilmore, L. A. and Price, N. C. (1995). Cloning and sequencing of a gene encoding pyruvate kinase from schizosaccharomyces pombe; implications for quaternary structure and regulation of the enzyme. *FEMS Microbiol. Lett.* 134: 221–226.

Nakaoka, H. and Wakamoto, Y. (2017). Aging, mortality, and the fast growth trade-off of schizosaccharomyces pombe. *PLoS biology* 15.

Nichols, R. J., Sen, S., Choo, Y. J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K. M., Lee, K. J., Wong, A. et al. (2011). Phenotypic landscape of a bacterial cell. *Cell* 144: 143–156.

O'Geen, H., Henry, I. M., Bhakta, M. S., Meckler, J. F. and Segal, D. J. (2015). A genome-wide analysis of cas9 binding specificity using chip-seq and targeted sequence capture. *Nucleic acids research* 43: 3389–3404.

Olin-Sandoval, V., Yu, J. S. L., Miller-Fleming, L., Alam, M. T., Kamrad, S., Correia-Melo, C., Haas, R., Segal, J., Navarro, D. A. P., Herrera-Dominguez, L. et al. (2019). Lysine harvesting is an antioxidant strategy and triggers underground polyamine metabolism. *Nature* 572: 249–253.

Otsu, N. (1979). A threshold selection method from Gray-Level histograms. *IEEE Transactions on Systems, Man, and Cybernetics* 9: 62–66.

Oustrin, M. L., Belenguer, P., Leroy, D., Hoffmann, I. and Ducommun, B. (1995). Effect of phenylarsine oxide on the fission yeast schizosaccharomyces pombe cell cycle. *Biochimie* 77: 279–287.

Paczia, N., Nilgen, A., Lehmann, T., Gätgens, J., Wiechert, W. and Noack, S. (2012). Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms. *Microbial cell factories* 11: 122.

Palková, Z. and Váchová, L. (2006). Life within a community: benefit to yeast long-term survival. *FEMS microbiology reviews* 30: 806–824.

Papagiannakis, A., Niebel, B., Wit, E. C. and Heinemann, M. (2017). Autonomous metabolic oscillations robustly gate the early and late cell cycle. *Molecular cell* 65: 285–295.

Parker, S., Fraczek, M. G., Wu, J., Shamsah, S., Manousaki, A., Dungrattanalert, K., de Almeida, R. A., Estrada-Rivadeneyra, D., Omara, W., Delneri, D. et al. (2017). A resource for functional profiling of noncoding rna in the yeast saccharomyces cerevisiae. *RNA* 23: 1166–1171.

Parker, S., Fraczek, M. G., Wu, J., Shamsah, S., Manousaki, A., Dungrattanalert, K., de Almeida, R. A., Invernizzi, E., Burgis, T., Omara, W. et al. (2018). Large-scale profiling of noncoding RNA function in yeast. *PLoS Genet.* 14: e1007253.

Pearce, A. K., Crimmins, K., Groussac, E., Hewlins, M. J., Dickinson, J. R., Francois, J., Booth, I. R. and Brown, A. J. (2001). Pyruvate kinase (pyk1) levels influence both the rate and direction of carbon flux in yeast under fermentative conditions. *Microbiology* 147: 391–401.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011). Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* 12: 2825–2830.

Petersen, J. and Russell, P. (2016). Growth and the environment of schizosaccharomyces pombe. *Cold Spring Harbor Protocols* 2016: pdb–top079764.

Pfeiffer, T. and Morley, A. (2014). An evolutionary perspective on the crabtree effect. *Front Mol Biosci* 1: 17.

Ponjavic, J., Ponting, C. P. and Lunter, G. (2007). Functionality or transcriptional noise? evidence for selection within long noncoding rnas. *Genome research* 17: 556–565.

Ponomarova, O., Gabrielli, N., Sévin, D. C., Mülleder, M., Zirngibl, K., Bulyha, K., Andrejev, S., Kafkia, E., Typas, A., Sauer, U. et al. (2017). Yeast creates a niche for symbiotic lactic acid bacteria through nitrogen overflow. *Cell Syst* 5: 345–357.e6.

Poyatos, J. F. (2020). Genetic buffering and potentiation in metabolism. *PLOS Computational Biology* 16: e1008185.

Rallis, C., Codlin, S. and Bähler, J. (2013). TORC1 signaling inhibition by rapamycin and caffeine affect lifespan, global gene expression, and cell proliferation of fission yeast. *Aging Cell* 12: 563–573.

Rallis, C., Mulleder, M., Smith, G., Au, Y. Z., Ralser, M. and Bahler, J. (2020). Amino acids whose intracellular levels change most during aging alter chronological lifespan of fission yeast. *BioRxiv* .

Ralser, M. (2018). An appeal to magic? the discovery of a non-enzymatic metabolism and its role in the origins of life. *Biochemical Journal* 475: 2577–2592.

Reimand, J., Kull, M., Peterson, H., Hansen, J. and Vilo, J. (2007). g: Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic acids research* 35: W193–W200.

Reuben, J. and Kayne, F. J. (1971). Thallium-205 nuclear magnetic resonance study of pyruvate kinase and its substrates. evidence for a substrate-induced conformational change. *J. Biol. Chem.* 246: 6227–6234.

Rhodes, N., Morris, C. N., Ainsworth, S. and Kinderlerer, J. (1986). The regulatory properties of yeast pyruvate kinase. effects of NH4+ and k+ concentrations. *Biochem. J* 234: 705–715.

Rich, P. (2003). The molecular machinery of keilin's respiratory chain. *Biochemical Society transactions* 31: 1095–105.

Risinger, A. L., Cain, N. E., Chen, E. J. and Kaiser, C. A. (2006). Activity-dependent reversible inactivation of the general amino acid permease. *Molecular biology of the cell* 17: 4411–4419.

Rodríguez-López, M., Cotobal, C., Fernández-Sánchez, O., Borbarán Bravo, N., Oktriani, R., Abendroth, H., Uka, D., Hoti, M., Wang, J., Zaratiegui, M. et al. (2016). A CRISPR/Cas9-based method and primer design tool for seamless genome editing in fission yeast. *Wellcome Open Res* 1: 19.

Ronne, H. (1995). Glucose repression in fungi. *Trends in Genetics* 11: 12–17.

Rosenberg, A., Ene, I. V., Bibi, M., Zakin, S., Segal, E. S., Ziv, N., Dahan, A. M., Colombo, A. L., Bennett, R. J. and Berman, J. (2018). Antifungal tolerance is a subpopulation effect distinct from resistance and is associated with persistent candidemia. *Nature communications* 9: 1–14.

Ruiz, S. J., van't Klooster, J. S., Bianchi, F. and Poolman, B. (2017). Growth inhibition by amino acids in saccharomyces cerevisiae. *bioRxiv* p. 222224.

Saint, M., Bertaux, F., Tang, W., Sun, X.-M., Game, L., Köferle, A., Bähler, J., Shahrezaei, V. and Marguerat, S. (2019). Single-cell imaging and rna sequencing reveal patterns of gene expression heterogeneity during fission yeast growth and adaptation. *Nature microbiology* 4: 480–491.

Sauro, H. M. (2017). Control and regulation of pathways via negative feedback. *Journal of The Royal Society Interface* 14: 20160848.

Schormann, N., Hayden, K. L., Lee, P., Banerjee, S. and Chattopadhyay, D. (2019). An overview of structure, function, and regulation of pyruvate kinases. *Protein Sci.* 28: 1771–1784.

Sen, P. and Orešič, M. (2019). Metabolic modeling of human gut microbiota on a genome scale: an overview. *Metabolites* 9: 22.

Shah, N. A., Laws, R. J., Wardman, B., Zhao, L. P. and Hartman, J. L., 4th (2007). Accurate, precise modeling of cell proliferation kinetics from time-lapse imaging and automated image analysis of agar yeast culture arrays. *BMC Syst. Biol.* 1: 3.

Shimizu, K. and Matsuoka, Y. (2019). Redox rebalance against genetic perturbations and modulation of central carbon metabolism by the oxidative stress regulation. *Biotechnology advances* p. 107441.

Slack, F. J. and Chinnaiyan, A. M. (2019). The role of non-coding rnas in oncology. *Cell* 179: 1033–1055.

Speed, D., Cai, N., UCLEB Consortium, Johnson, M. R., Nejentsev, S. and Balding, D. J. (2017). Reevaluation of SNP heritability in complex human traits. *Nat. Genet.* 49: 986–992.

Speed, D., Hemani, G., Johnson, M. R. and Balding, D. J. (2012). Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.* 91: 1011–1021.

Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Botstein, D. and Futcher, B. (1998). Comprehensive identification of cell cycle–regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. *Molecular biology of the cell* 9: 3273–3297.

Stevenson, K., McVey, A. F., Clark, I. B., Swain, P. S. and Pilizota, T. (2016). General calibration of microbial growth in microplate readers. *Scientific reports* 6: 1–7.

Sun, F., Dai, C., Xie, J. and Hu, X. (2012). Biochemical issues in estimation of cytosolic free NAD/NADH ratio. *PLoS One* 7: e34525.

Szenk, M., Dill, K. A. and de Graff, A. M. R. (2017). Why do Fast-Growing bacteria enter overflow metabolism? testing the membrane real estate hypothesis. *Cell Syst* 5: 95–104.

Takeda, K., Starzynski, C., Mori, A. and Yanagida, M. (2015). The critical glucose concentration for respiration-independent proliferation of fission yeast, schizosaccharomyces pombe. *Mitochondrion* 22: 91–95.

Takeuchi, R., Tamura, T., Nakayashiki, T., Tanaka, Y., Muto, A., Wanner, B. L. and Mori, H. (2014). Colony-live–a high-throughput method for measuring microbial colony growth kinetics–reveals diverse growth effects of gene knockouts in escherichia coli. *BMC Microbiol.* 14: 171.

Thommes, M., Wang, T., Zhao, Q., Paschalidis, I. C. and Segrè, D. (2019). Designing metabolic division of labor in microbial communities. *Msystems* 4.

Till, P., Mach, R. L. and Mach-Aigner, A. R. (2018). A current view on long noncoding rnas in yeast and filamentous fungi. *Applied microbiology and biotechnology* 102: 7319–7331.

Traven, A., Jänicke, A., Harrison, P., Swaminathan, A., Seemann, T. and Beilharz, T. H. (2012). Transcriptional profiling of a yeast colony provides new insight into the heterogeneity of multicellular fungal communities. *PLoS One* 7.

Trcek, T., Chao, J. A., Larson, D. R., Park, H. Y., Zenklusen, D., Shenoy, S. M. and Singer, R. H. (2012). Single-mrna counting using fluorescent in situ hybridization in budding yeast. *Nature protocols* 7: 408.

Tsukada, M. and Ohsumi, Y. (1993). Isolation and characterization of autophagy-defective mutants of saccharomyces cerevisiae. *FEBS Lett.* 333: 169–174.

Uygun, S., Peng, C., Lehti-Shiu, M. D., Last, R. L. and Shiu, S.-H. (2016). Utility and limitations of using gene expression data to identify functional associations. *PLoS computational biology* 12: e1005244.

Váchová, L., Hatáková, L., Čáp, M., Pokorná, M. and Palková, Z. (2013). Rapidly developing yeast microcolonies differentiate in a similar way to aging giant colonies. *Oxidative medicine and cellular longevity* 2013.

Váchová, L., Kučerová, H., Devaux, F., Úlehlová, M. and Palková, Z. (2009). Metabolic diversification of cells during the development of yeast colonies. *Environmental microbiology* 11: 494–504.

Váchová, L. and Palková, Z. (2018). How structured yeast multicellular communities live, age and die? *FEMS Yeast Res.* 18.

Valvezan, A. J. and Manning, B. D. (2019). Molecular logic of mTORC1 signalling as a metabolic rheostat. *Nature Metabolism* 1: 321–333.

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., Yu, T. and scikit-image contributors (2014). scikit-image: image processing in python. *PeerJ* 2: e453.

Varahan, S., Sinha, V., Walvekar, A., Krishna, S. and Laxman, S. (2020). Resource plasticity-driven carbon-nitrogen budgeting enables specialization and division of labor in a clonal community. *eLife* 9: e57609.

Varahan, S., Walvekar, A., Sinha, V., Krishna, S. and Laxman, S. (2019). Metabolic constraints drive self-organization of specialized cell groups. *Elife* 8: e46735.

Vassiliadis, D., Wong, K. H., Andrianopoulos, A. and Monahan, B. J. (2019). A genome-wide analysis of carbon catabolite repression in schizosaccharomyces pombe. *BMC genomics* 20: 251.

Vazquez, A. and Oltvai, Z. N. (2016). Macromolecular crowding explains overflow metabolism in cells. *Sci. Rep.* 6: 31007.

Veal, E. A., Tomalin, L. E., Morgan, B. A. and Day, A. M. (2014). The fission yeast schizosaccharomyces pombe as a model to understand how peroxiredoxins influence cell responses to hydrogen peroxide. *Biochem. Soc. Trans.* 42: 909–916.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J. et al. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17: 261–272.

Vivancos, A. P., Jara, M., Zuin, A., Sansó, M. and Hidalgo, E. (2006). Oxidative stress in schizosaccharomyces pombe: different H2O2 levels, different response pathways. *Mol. Genet. Genomics* 276: 495–502.

Vowinckel, J., Zelezniak, A., Bruderer, R., Mülleder, M., Reiter, L. and Ralser, M. (2018). Cost-effective generation of precise label-free quantitative proteomes in high-throughput by microlc and data-independent acquisition. *Scientific reports* 8: 1–10.

Wagih, O. and Parts, L. (2014). gitter: a robust and accurate method for quantification of colony sizes from plate images. *G3* 4: 547–552.

Warburg, O. (1927). The metabolism of tumors in the body. *The Journal of General Physiology* 8: 519–530.

Wellen, K. E., Lu, C., Mancuso, A., Lemons, J. M., Ryczko, M., Dennis, J. W., Rabinowitz, J. D., Coller, H. A. and Thompson, C. B. (2010). The hexosamine biosynthetic pathway couples growth factor-induced glutamine uptake to glucose metabolism. *Genes & development* 24: 2784–2799.

Wery, M., Gautier, C., Descrimes, M., Yoda, M., Vennin-Rendos, H., Migeot, V., Gautheret, D., Hermand, D. and Morillon, A. (2018). Native elongating transcript sequencing reveals global anti-correlation between sense and antisense nascent transcription in fission yeast. *Rna* 24: 196–208.

Wickham, H. et al. (2014). Tidy data. *Journal of Statistical Software* 59: 1–23.

Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S. et al. (2002). The genome sequence of schizosaccharomyces pombe. *Nature* 415: 871–880.

Wood, V., Harris, M. A., McDowall, M. D., Rutherford, K., Vaughan, B. W., Staines, D. M., Aslett, M., Lock, A., Bähler, J., Kersey, P. J. et al. (2012). PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Res.* 40: D695–9.

Wood, V., Lock, A., Harris, M. A., Rutherford, K., Bähler, J. and Oliver, S. G. (2019). Hidden in plain sight: what remains to be discovered in the eukaryotic proteome? *Open biology* 9: 180241.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W. and Steinmetz, L. M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457: 1033–1037.

Yamada, E. A. and Sgarbieri, V. C. (2005). Yeast (saccharomyces cerevisiae) protein concentrate: preparation, chemical composition, and nutritional and functional properties. *Journal of Agricultural and food chemistry* 53: 3931–3936.

Yu, T., Zhou, Y. J., Huang, M., Liu, Q., Pereira, R., David, F. and Nielsen, J. (2018). Reprogramming yeast metabolism from alcoholic fermentation to lipogenesis. *Cell* 174: 1549–1558.e14.

Zackrisson, M., Hallin, J., Ottosson, L.-G., Dahl, P., Fernandez-Parada, E., Ländström, E., Fernandez-Ricaud, L., Kaferle, P., Skyman, A., Stenberg, S. et al. (2016). Scan-o-matic: High-Resolution microbial phenomics at a massive scale. *G3* 6: 3003–3014.

Zaman, S. B., Hussain, M. A., Nye, R., Mehta, V., Mamun, K. T. and Hossain, N. (2017). A review on antibiotic resistance: alarm bells are ringing. *Cureus* 9.

Zelezniak, A., Andrejev, S., Ponomarova, O., Mende, D. R., Bork, P. and Patil, K. R. (2015). Metabolic dependencies drive species co-occurrence in diverse microbial communities. *Proceedings of the National Academy of Sciences* 112: 6449–6454.

Zelezniak, A., Vowinckel, J., Capuano, F., Messner, C. B., Demichev, V., Polowsky, N., Mülleder, M., Kamrad, S., Klaus, B., Keller, M. A. et al. (2018). Machine learning predicts the yeast metabolome from the quantitative proteome of kinase knockouts. *Cell Systems* 7: 269–283.

Zhang, W., Du, G., Zhou, J. and Chen, J. (2018). Regulation of sensing, transportation, and catabolism of nitrogen sources in saccharomyces cerevisiae. *Microbiology and Molecular Biology Reviews* 82.

Zhuang, K., Vemuri, G. N. and Mahadevan, R. (2011). Economics of membrane occupancy and respiro-fermentation. *Mol. Syst. Biol.* 7: 500.

Zuk, O., Hechter, E., Sunyaev, S. R. and Lander, E. S. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences* 109: 1193–1198.

Zwietering, M., Jongenburger, I., Rombouts, F. and Van't Riet, K. (1990). Modeling of the bacterial growth curve. *Appl. Environ. Microbiol.* 56: 1875–1881.