CHALMERS
UNIVERSITY OF TECHNOLOGY

Exploring the mycobiota for the treatment of gut-related diseases

Master's thesis in Biotechnology

ANDREA CLAUSEN LIND

Supervisor: Verena Siewers
Co-supervisor: Raphael Ferreira

DEPARTMENT OF BIOLOGY AND BIOLOGICAL ENGINEERING
DIVSION OF SYSTEMS AND SYNTHETIC BIOLOGY

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020
www.chalmers.se

Abstract

Our gut harbors trillions of microorganisms that form a bridge between our diet and whole-body metabolism. This collection of microorganisms, referred to as the gut microbiota, can in some cases be perturbed and lead to the onset of disease. Studies exploring the gut microbiota have identified a plethora of bacteria implicated in multiple diseases; however, the fungal component of the microbiome, known as the mycobiome, remains largely unexplored. Similarly, engineered live biotherapeutics designed to target these diseases have mainly been limited to bacterial chassis. This makes the mycobiome a promising target for exploratory and therapeutic efforts in the understanding and management of gut related diseases. Herein, we aimed to engineer the probiotic yeast *Saccharomyces boulardii* as a live biotherapeutic for treatment of the inborn metabolic disease phenylketonuria by making the yeast consume large amounts of phenylalanine. An improved consumption of phenylalanine was observed in several of the resulting stains compared to control strains, but the overall consumption was still far from what would be required for effective treatment of the disease. As an offshoot of this project, we also aimed to develop and apply an effective platform for analysis of the fungal composition in metagenomic data. This was achieved by modifying an existing fungal metagenomics pipeline and subsequently analyzing the mycobiome of ~1300 fecal samples from patients with Inflammatory Bowel Disease (IBD), obtained from the integrative Human Microbiome Project data set. Several fungi were found to be slightly but significantly enriched in IBD patients compared to healthy controls. Furthermore, some of the identified fungi have been identified in previous studies of the mycobiome in IBD, validating our approach. In summary, this thesis highlights the unexplored potential of the human gut mycobiome across diseases.

Keywords: Yeast Probiotics, Live biotherapeutics, Phenylketonuria, Mycobiome, Metagenomic analysis, Inflammatory Bowel Disease

Acknowledgements

I am happy to have done my one-year long master's theses at SysBio, as I have gotten the chance to meet and get to know so many fun, intelligent, and passionate people along the way. I am especially thankful to my supervisor Verena Siewers and co-supervisor Raphael Ferreira for giving me the opportunity to work on this project, and for the constant support and advice you both have given along the way. I have learned so much from both of you, and you are both a great inspiration as scientists.

I also want to say thank you to Angelo Limeta for collaborating with me on the metagenomic analysis of the mycobiome, both for being an inspiration and for making the work interesting and enjoyable.

Moreover, I want to thank David Lund for being a great opponent and for giving valuable feedback on my thesis. I also want to thank you and everyone we studied with for making studying at Chalmers fun and memorable.

For helping me in this project I want to thank Paul Donovan, who made the original FindFungi and who helped with further development of the pipeline through advice, as well as Alex Hedin who gave advice on optimization of transformation in *S. boulardii*. Thank you also to Rui Pereira for advising me on the application of the ALE, Maximillian Otto for advising me on construction of cloning vectors, and Olena Ishchuk and John Hellgren for advising me on ways to increase transformation efficiency.

I also want to thank the laboratory engineers and admins for keeping all of SysBio running, and everyone else working in the lab and in the Dungeon for always answering questions and helping me whenever I was lost while making the work fun and exciting. Furthermore, I want to thank everyone at CMSI for all the help I got with running the LC-MS/MS analysis, would not have been able to do it without your guidance.

Lastly, I want to extend a special thanks to Angelo Limeta and to my family and friends for always supporting me.

Contents

1 Introduction

The gut microbiota encompasses the total set of microorganisms present in the gut, which includes bacteria, fungi, archaea, protists, and viruses. Bacteria were first observed in fecal material during the 1700s, but could not be properly studied further until the mid-1900s when methods for anaerobic culturing was developed for the first time, enabling cultivation of previously undetectable organisms (Clark, 2019; Pariente, 2019). The first fecal transplantation was successfully performed around this time for treatment of *Clostridium difficile* infection, resulting in rapid recovery for all four patients, thereby demonstrating the importance of the gut microbiota composition for health (Stone, 2019). Since then the field has rapidly expanded, largely driven by the development of new tools (Pace, 1997; Pariente, 2019). These include germ-free animals, which allow researchers to obtain a mechanistic understanding of individual host-microbe interactions; as well as high throughput sequencing technologies, which allow for the identification of previously unculturable microbes (Brunello, 2019; Pariente, 2019; Tang, 2019). These techniques have now allowed us to map out specific gut microbiota profiles linked to multiple diseases, including type 2 diabetes, cancer, Infalmmatory Bowel Disease (IBD), depression, and obesity (Qin et al., 2012; Rooks et al., 2014; Valles-Colomer et al., 2019; Wang and Jia, 2016; Zmora et al., 2019).

The increased knowledge gathered on the health impact of different bacteria has prompted interest in using bacteria as therapeutic vessels. Examples of engineered bacteria as targeted disease treatments include *Lactococcus lactis* engineered to locally combat colitis, *Salmonella thyphimurium* engineered to target and treat tumors through subcutaneous injection, *Escherichia coli* engineered to treat symptoms of metabolic disease from the gut (Isabella et al., 2018; Ozdemir et al., 2018). While research in the area of bacteria as live biotherapeutics is substantial, there are comparably fewer attempts of engineering yeast for similar purposes (Hudson et al., 2016; Sabu et al., 2019). The potential of yeasts as engineered live biotherapeutics has been overlooked until recently, the fungal component of the gut microbiota, coined the mycobiome, is only beginning to be explored. A reason for this lack of interest has been that the bacterial component of the microbiota comprises approximately 99% of all microbes present in the gut (Qin et al., 2010). This makes the bacterial component a more likely source of influence to the host. Additionally, the higher relative abundance in the gut makes it easier to perform quantitative analysis on the bacterial reads using shotgun metagenomic sequencing (Richard and Sokol, 2019). Nonetheless, research conducted in recent years has established the mycobiome as an emerging player across several diseases, which has renewed interests in the field. Examples include altered mycobiome composition in patients with IBD compared to healthy individuals (Mukherjee et al., 2015; Richard and Sokol, 2019). Additionally, a landmark study by Aykut and colleagues reported that fungi belonging to the *Malassezia* genus causally promote the initiation and development of pancreatic cancer in mice, by translocating from the duodenum in the gut to the pancreatic duct via the sphincter of Oddi (Aykut et al., 2019). These findings highlight the future potential of engineering and exploiting host-mycobiome interactions for the treatment of disease.

In this study, we aimed to engineer the probiotic yeast *Saccharomyces boulardii* for increased consumption of the amino acid phenylalanine, with the aim of developing a yeast-based live biotherapeutic for treatment of the inborn metabolic disease phenylketonuria (PKU). As an offshoot, we also aimed to develop a computational pipeline for quantitative analysis of fungal composition in metagenomic data. The pipeline will then be applied to the IBD dataset from the integrative Human Microbiome Project (iHMP) database, in order to examine if it is possible to discern quantitative differences in fungal composition of the microbiota between

healthy subjects and subjects with IBD using metagenomic data. Overall, we hope to establish yeast as a promising live biotherapeutic, further elucidate the involvement of the mycobiome in health and disease, as well as create a platform for efficient analysis of the fungal composition in metagenomic datasets.

2 Theory

2.1 Diseases of metabolism

A collection of diseases to which an engineered live biotherapeutic is applicable are metabolic diseases. These are caused by inborn genetic mutations, resulting in decreased or non-existent activity of enzymes involved in metabolism and consequentially causing metabolites involved in the affected pathway to either accumulate or become scarce in the body (Rice and Steiner, 2016). Depending on the metabolic pathway involved, the symptoms of disease differ from physical unease, physical or congenital disability, or even death (Nyhan, W.L. et al., 2012). Current therapies for these diseases only aim to alleviate symptoms, as reliable cures for diseases caused on a genetic level are yet to be developed (Rice and Steiner, 2016).

2.1.1 Phenylketonuria

Phenylketonuria (PKU) is an autosomal recessive metabolic disease affecting phenylalanine metabolism, resulting in accumulation of the amino acid phenylalanine (Phe) in the blood and brain (van Wegberg et al., 2017). The disease is caused by dysfunction of the enzyme phenylalanine hydroxylase (PAH), originating from genetic mutations or from a lack of co-enzymes, which catalyzes conversion of Phe into tyrosine in cells of the human body (Fig. 1; van Wegberg et al., 2017; Williams et al., 2008). While the blood concentration of Phe in a healthy subject lies below 400 µM, concentrations range from 600 µM to over 1200 µM in a PKU subject depending of the severity of dysfunction in PAH (Hanley, 2004). Formally two severity levels of PKU are defined based on Phe blood concentrations; classic PKU (Phe > 1200 µM) and mild PKU (600 µM < Phe < 1200 µM), while Phe levels between 400 µM and 600 µM are defined as mild hyperphenylalaninemia (HPA) (Hanley, 2004). If not treated immediately after birth the high blood and brain levels of Phe can result in severe intellectual disability, microcephaly, aggressive behavior, seizures, and anxiety in affected subjects (“Phenylketonuria. Genetic and Rare Diseases Information Center (GARD)”; van Wegberg et al., 2017). The prevalence of PKU varies worldwide, affecting 1-5 in 10000 newborns every year, being particularly prevalent in Turkey where 1 in 4000 newborns is affected yearly (“Orphanet: Phenylketonuria”). The toxicity pf PKU is thought to result from Phe-mediated saturation of amino acid transporters shared by Phe and other bulky amino acids, caused by the relative abundance of Phe over other amino acids in the blood (Pietz et al., 1999). The consequent disruption of transport of other bulky amino acids into the cells results in a shortage of building blocks for new proteins required for cellular maintenance and proliferation (Vliet et al., 2015).

Phe is an essential amino acid, meaning it cannot be synthesized *de novo* in the body but has to be ingested through proteins in the diet (van Spronsen, 2010). This also means that Phe accumulation in PKU subjects can be prevented through a Phe restricted diet, which has been the most prevalent form of treatment to this day (“Phenylketonuria. Genetic and Rare Diseases Information Center (GARD)”; van Wegberg et al., 2017). Newborn screening programs for PKU were later developed and implemented in most countries, enabling early detection and intervention to prevent progression of disease symptoms by introducing a strict Phe-regulated diet from the first week of birth (”Phenylketonuria. Genetic and Rare Diseases Information

Center (GARD)"; van Wegberg et al., 2017). This solution remains suboptimal, as it has been observed that small deviances are common from the lifelong Phe-regulated diet, resulting in proportional cognitive

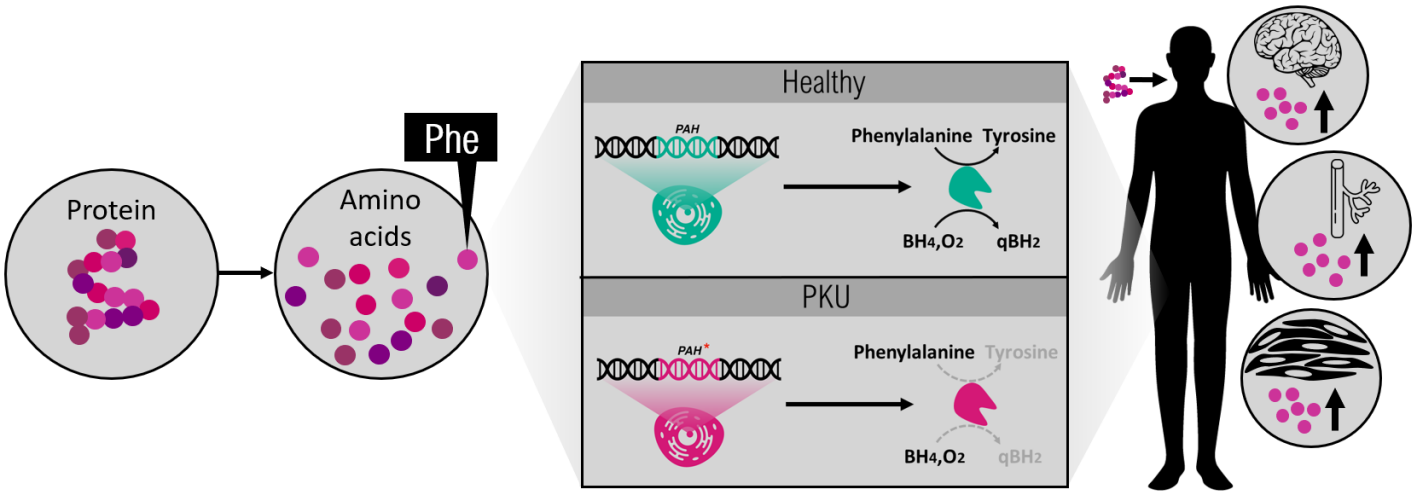

Figure 1. Phenylalanine metabolism in *H. sapiens*. Proteins are composed of amino acids, including Phe. In a healthy individual (top middle) PAH together with co-enzyme BH_4 and molecular oxygen catalyzes the reaction converting phenylalanine into tyrosine. In an individual with PKU (bottom middle) inborn genetic mutations in the gene encoding PAH cause decreased to non-existent function of the enzyme, preventing conversion of Phe into tyrosine and causing subsequent accumulation of Phe. In a PKU patient, ingestion of protein causes elevated brain, blood, and tissue levels of Phe.

decline (Al Hafid and Christodoulou, 2015). The cause of poor dietary compliance has been attributed to the lack of palatable options within the limits of the diet, consisting of low protein foods together with supplements to avoid nutritional deficiency of other amino acids and as well as vitamins B12 and D (Al Hafid and Christodoulou, 2015). Dietary non-compliance is especially common in children approaching adolescence, resulting in lifelong effects (Al Hafid and Christodoulou, 2015).

Other than dietary restriction, Sapropterin and Pegvaliase have been approved by the U.S. Food and Drug Administration (FDA) for treatment of PKU and HPA ("Phenylketonuria. Genetic and Rare Diseases Information Center (GARD)"). Sapropterin, also known as tetrahydrobiopterin or BH_4 which is the co-enzyme of PAH, is only applicable in about 30% of all cases where patients are BH_4-responsive and still has to be accompanied by a Phe-restricted diet ("Phenylketonuria. Genetic and Rare Diseases Information Center (GARD)"; van Wegberg et al., 2017). Pegvaliase is an enzyme-based therapy where patients receive injections of PEGylated phenylalnine-ammonia lyase (PAL), converting phenylalanine into the non-toxic flavonoid trans-cinnamic acid (tCA) (Al Hafid and Christodoulou, 2015; van Wegberg et al., 2017). Although effective, this treatment is currently only given to patients with the most severe forms of PKU, where Phe blood levels exceed 600 µM even with a strict Phe-regulated diet, due to the risk of severe and potentially life-threatening allergic reactions that may occur any time during treatment ("PALYNZIQ® (pegvaliase-pqpz) Injection for PKU"; "Phenylketonuria. Genetic and Rare Diseases Information Center (GARD)"). It has also been proposed that Pegvaliase could be used as a means for patients to consume a less strict, or even an unregulated, diet while still maintaining normal blood Phe levels ("Powerful PKU control without a low-Phe diet requirement*"). The treatment does however require weekly injections which could pose a burden on families, on top of a financial burden as the treatment reportedly costs 192,000 euros yearly in the US (Deena, 2018). This is compared to the cost of supplements and foods needed in a Phe-regulated diet, estimated 4,600 euros yearly in the UK (Al Hafid and

Christodoulou, 2015). Oral Pegvaliase treatment has been attempted, but the enzyme proved to be less stable in the intestinal tract making injections the continued way of delivery (Al Hafid and Christodoulou, 2015). Few can afford a lifelong Pegvaliase treatment, and together with the risk of potential immunological adverse effects it becomes clear that a less invasive, cheaper, and safer alternative is needed.

A live biotherapeutic engineered to convert large amounts of phenylalanine into other, harmless, compounds in the digestive tract could be a viable alternative. Potentially, it could prove to be a cheaper alternative to enzyme therapy, considering that production and purification of enzymes are more expensive compared to that of a microorganism. Using a live biotherapeutic vessel also ensures stability of enzymes the system, as all enzymes produced for Phe consumption will be protected inside the cell, removing the need for injections while potentially maintaining the benefits of a less strict diet for patients. The bacterium *Escherichia coli Nissle* has already been engineered for treatment of PKU and is currently in phase 2 clinical trials("Safety and Tolerability of SYNB1618 in Healthy Adult Volunteers and Adult Subjects With Phenylketonuria"). However, yeast has on several other occasions proven to be an alternative, or even preferred, vessel for biotechnological applications such as insulin and fatty acid production (Baeshen et al., 2014; Gajewski et al., 2017). It is therefore relevant to explore yeast, in this case the probiotic yeast *Saccharomyces boulardii*, as an alternative live biotherapeutic vessel for treatment of PKU.

2.2 Inflammatory diseases of the gut

IBD includes multiple inflammatory diseases of the gut, of which Ulcerative colitis (UC) and Crohn's disease (CD) are the most common ones. Symptoms include pain, bloating, an erratic bowel habit, and in the acute phase bloody diarrhea may occur (Spiller and Major, 2016). Due to the cause of the disease being unknown, diagnosis largely relies on chronic remissions and relapses (Spiller and Major, 2016). Several factors affect the onset of the disease, including psychological stressors, an impaired gut-barrier function, altered enteric nerves, and dysbiosis of the gut microbiota (Spiller and Major, 2016). The bacterial microbiota in patients with IBD has been thoroughly investigated, but only a few studies have focused on the mycobiota composition (Lewis et al., 2015; Miyoshi et al., 2018; Sokol et al., 2017). Several fungi have been found to be associated with the disease, although small cohort sizes have limited any general conclusions.

2.3 *Saccharomyces boulardii as a* live biotherapeutic

The probiotic yeast *S. boulardii*, a close relative to baker's yeast *Saccharomyces cerevisiae*, has long been used in treatment of digestive disorders (Kelesidis and Pothoulakis, 2012). It was first isolated from lychee and mangosteen skins in 1923 and even before that the native Indonesian population chewed on lychee and mangosteen skins to prevent the same diseases (Altmann, 2017). It is now sold as a yeast probiotic, shown to promote homeostasis and recovery of the gut microbiota after antibiotic treatment in young and adults, strengthen the gut-lumen barrier, prevent mucosal colonization of pathogenic bacteria, and reduce activity of inflammatory pathways mucosal cells (Agamennone et al., 2018; Kelesidis and Pothoulakis, 2012; Moré and Swidsinski, 2015). Due to its long history in medical care, *S. boulardii* is considered safe for consumption and has a Generally Regarded As Safe (GRAS) status, making the yeast an ideal platform for further development of a biotherapeutic vessel (Edwards-Ingram et al., 2007). Moreover, other desirable characteristics such as tolerance of the varying pH levels of the digestive system, inherent ability to fold complex proteins and make additional post-translational protein modifications not found in bacteria, makes it a suitable live biotherapeutic

vessel (Edwards-Ingram et al., 2007; Hudson et al., 2016). The similarity of the *S. boulardii* and the well characterized *S. cerevisiae* is another advantage, as engineering methods are well established for the latter. *S. boulardii* cannot sporulate however, meaning that all strains are diploid and carry two copies of each gene (Edwards-Ingram et al., 2007). Recently a study preprint by Deniz Durmusoglu and colleagues was posted on bioRxiv exploring the use of *S. cerevisiae* promoters, terminators, plasmids, and marker genes in *S. boulardii*, overall establishing compatibility (Durmusoglu et al., 2020). Furthermore, the yeasts share >99% of their genomic contents, making it possible to apply software developed for identification of genomic targets in *S. cerevisiae* to *S. boulardii* as well (Khatri et al., 2017). These factors make the yeast *S. boulardii* a promising live biotherapeutic vessel for application in the treatment of diseases across all parts of the digestive tract.

2.3.1 Engineering strategy of *S. boulardii*

In order to increase consumption of Phe in *S. boulardii* a combination of targeted engineering and evolutionary techniques was applied. Using a targeted engineering approach based on metabolic pathways in *S. cerevisiae*, together with introduction of heterologous proteins in the yeast, we set out to rewire the native Phe metabolism of *S. boulardii*. By in parallel running an evolutionary experiment where we aimed to evolve a yeast for increased Phe consumption or to possibly grow solely on phenylalanine, similar to the yeast *Rhodotorula glutinis*, we hoped to discover novel engineering targets for our yeast (Marusich et al., 1981). Phenylalanine synthesis in *S. boulardii* begins in glycolysis where glucose-6-phosphate enters the shikimate pathway, in which it is metabolized into chorismate, a substrate for several flavonoids as well as amino acids Phe and tyrosine (Tyr) (**Fig. 2)**. Phe and Tyr are produced through the arogenate pathway, where several enzymes are involved in synthesis of both amino acids. Phe is not only produced *de novo* in the cells but is also taken up from the environment, primarily through the transporters Agp1p and Gap1p, but also through to some extent through the transporters Bap1p, Bap2p, and Tat2p (Regenberg and Kielland-Brandt, 1999; Sáenz et al., 2014).

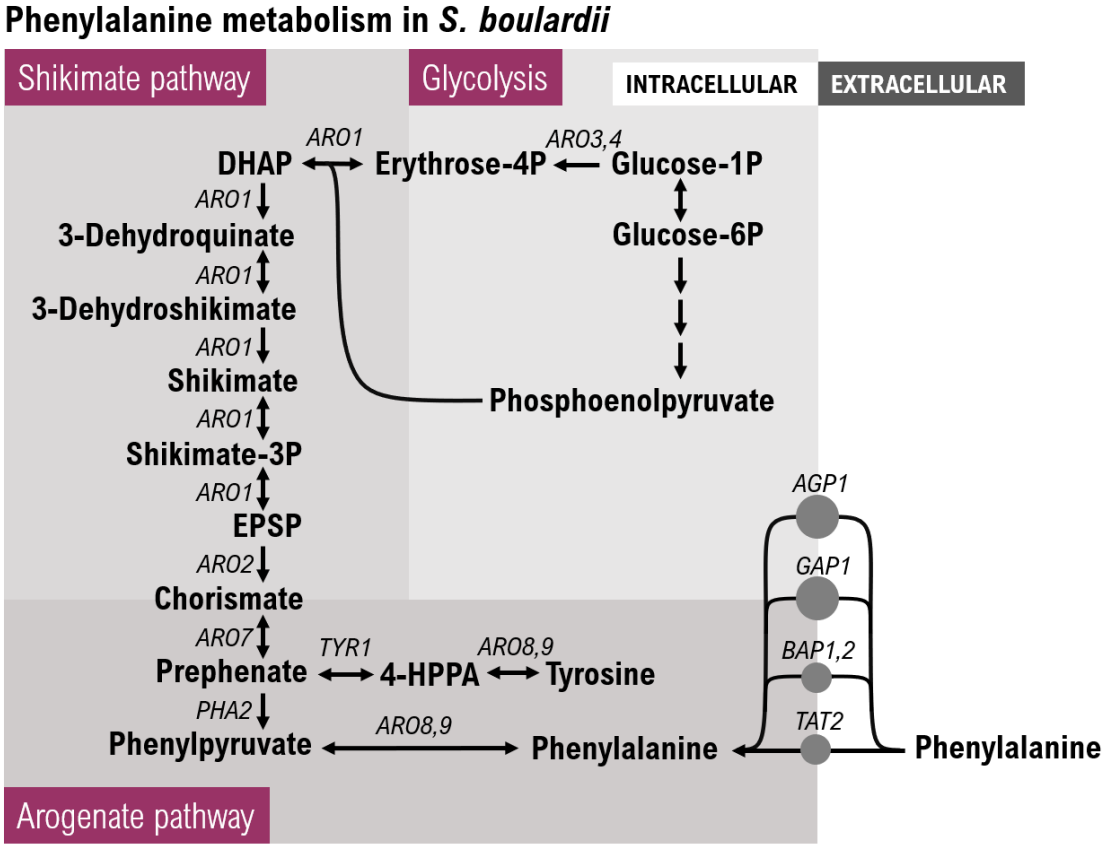

Figure 2. Phenylalanine metabolism in *S. boulardii*, starting from glycolysis, going through the shikimate pathway, and ending in the arogenate pathway. Genes encoding enzymes and transporters responsible for different steps of the illustrated pathways are shown in italic. Metabolites are shown in bold. Glucose-1P: Glucose-1-phosphate, Glucose-6P: Glucose-6-phosphate, Erythrose-4P: Erythrose-4-phosphate, DHAP: 3-Deoxy-D-arabinoheptulosonate 7-phosphate, Shikimate-3P: Shikimate-3-phosphate, EPSP: 5-Enolpyruvylshikimate-3-phosphate, 4-HPPA: 4-Hydroxyphenylpyruvic acid.

2.3.1.1 Metabolic engineering

Two engineering approaches were designed for increased consumption of Phe in *S. boulardii* (**Fig. 3)**. In the primary engineering approach, *ARO8* and *ARO9* were to be deleted in order to make the strains Phe auxotrophic. As this would also result in Tyr auxotrophy, *PAH* from *Homo sapiens* was to be inserted in the genomic locus of *ARO8* under expression of constitutive promoter P_{TEF1}. As PAH catalyzes conversion of Phe to Tyr, the resulting strain would no longer be Tyr auxotrophic and Phe consumption would be increased. The gene encoding PAL from *Rhodosporidium toruloides*, converting Phe into trans-cinnamate (tCA), would be inserted in the genomic locus of *ARO9* under P_{TEF1} control which would again increase Phe consumption. Furthermore, as proteins expressed for Phe consumption were intracellular, transport of amino acid into the cell had to be increased for the total consumption to be enhanced. As no yeast has amino acid transporters specific to Phe, the native *AGP1* transporter gene was to be overexpressed under P_{TEF1}. This is due to Agp1p being one of the primary Phe transporters in *S. boulardii*.

The second engineering approach was largely the same as the previous one, except that we instead deleted *PHA2* to induce auxotrophy. *PAL* would then be inserted in the *PHA2* genomic locus instead, which would result in a strain with Phe auxotrophy.

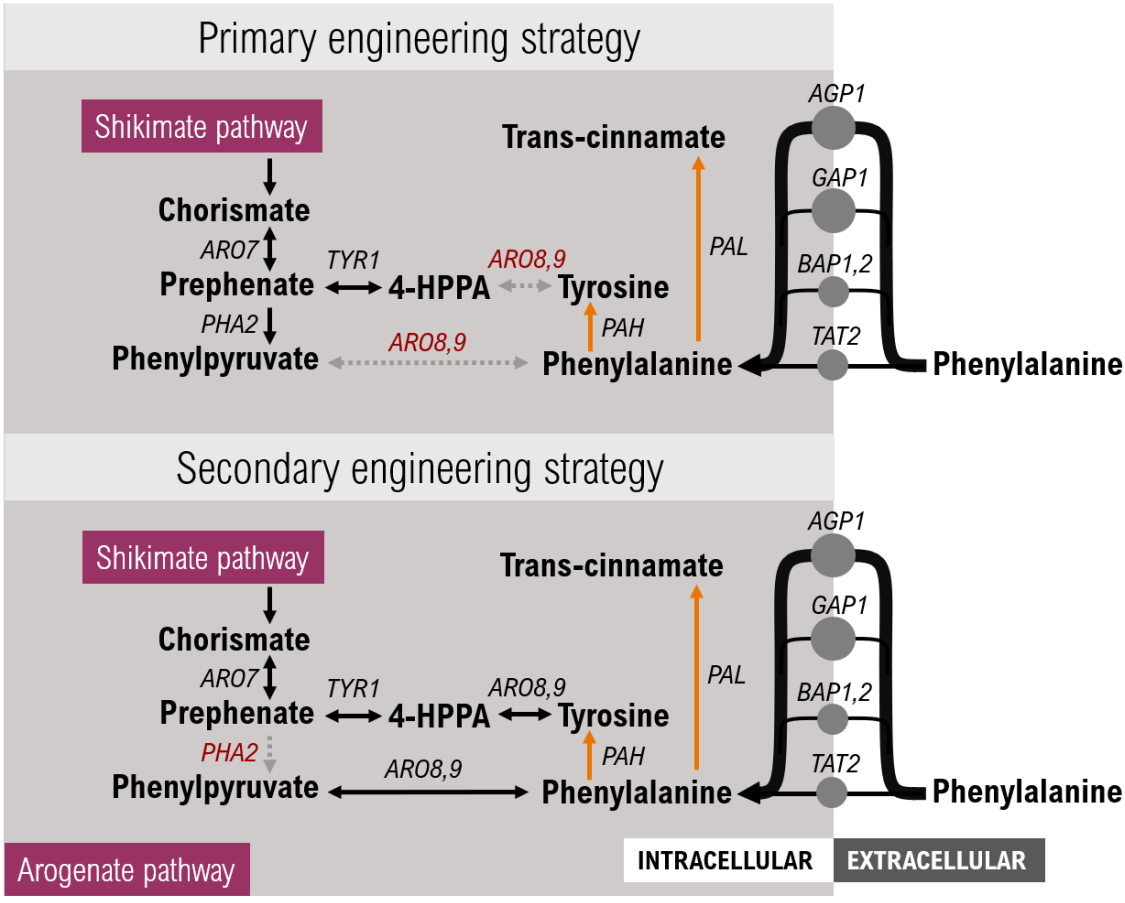

Figure 3. Engineering approaches to obtain a *S. boulardii* strain with high consumption of Phe. The primary strategy (top) includes deletion of genes *ARO8* and *ARO9* to create a Phe and Tyr auxotrophic strain, introduction of the heterologous PAH for conversion of Phe to Tyr, introduction of the heterologous protein PAL for conversion of Phe to tCA, and overexpression of amino acid transporter Agp1p. The secondary engineering strategy (bottom) is largely the same with the exception of *PHA2* being deleted for auxotrophy instead of *ARO8* and *ARO9*. 4-HPPA: 4-Hydroxyphenylpyruvic acid.

2.3.1.2 Evolutionary Engineering

Although targeted engineering of strains is an effective approach to increase Phe consumption, it is limited to known targets. In Adaptive Laboratory Engineering (ALE), the aim is either to improving existing characteristics of the organism or to evolve new functions such as growth on a new carbon source (Dragosits and Mattanovich, 2013). This is possible through application of a defined cultivation condition designed to create a selective pressure in the cellular cultures, benefitting growth of cells that develop the desired phenotype (Dragosits and Mattanovich, 2013). The possibility of such a phenotype occurring increases with the time of cultivation, as the likelihood of a beneficial mutation taking place increases with time. The evolutionary pressure is typically designed to confer a growth advantage to individual cells in the culture that obtain a genomic mutation conferring the desired trait. Thus, evolved cells will outgrow other cells in the culture over time, until a new beneficial mutation occurs, and the new strain becomes dominant in the culture. In order to discern the mechanisms behind the evolved traits, the strains with promising characteristics are sequenced and genomes are compared to that of the original strain. Reverse engineering is often applied to confirm which mutation resulted in the observed trait, allowing for identification of the mechanisms underlying the improved strain characteristics. As an example, this approach was successfully applied for identification of genetic targets promoting conversion of 5-hydroxymethyl furfural, an inhibitor of ethanol production found in hydrolyzed lignocellulosic materials, into 5-hydroxymethylfurfural alcohol (Petersson et al., 2006). Through microarray expression analysis of two strains with differentiating abilities to convert 5-hydroxymethyl furfural *ADE6* was found to have a higher expression level in the better performing strain (Petersson et al., 2006). By later overexpressing *ADE6* in a wild type strain the significance of the gene in conversion of 5-hydroxymethyl furfural to 5-hydroxymethylfurfural alcohol was confirmed (Petersson et al., 2006).

In this project the aim was to either evolve a strain that consumes Phe at a higher rate, or that can grow solely on Phe as a carbon and nitrogen source similar to the red yeast *Rhodorula glutinis* (Marusich et al., 1981).

2.4 The human gut mycobiome

Having discussed yeast as potential treatment vessels for treatment of metabolic disease, it is also relevant to bring up fungi as potential factors promoting or preventing disease in the gut microbiota. Although the abundance of fungi is small compared to that of bacteria in the gut microbiota, with 10^{5}-10^{6} times larger abundance of bacteria per gram feces, it has been shown that the kingdoms interact and that both are essential for microbiota homeostasis (Richard and Sokol, 2019; Sam et al., 2017). Just as the presence of bacteria has been shown to prevent overgrowth and pathogenesis of certain fungi, the presence of fungi has been shown to prevent overgrowth of bacteria in mice (Sam et al., 2017). In the study, bacteria that are normally considered as contributors to a healthy gut environment were shown to overgrow and induce inflammation in mice treated with anti-fungals (Sam et al., 2017). Further taking into consideration the difference in size between fungi and bacteria, with fungi being about 100-fold larger in volume, it is plausible that fungi contribute to a larger extent than can be estimated by abundance alone (Sam et al., 2017). The low fungal abundance did however cause the fungal

composition to be deemed irrelevant until recently. Since the mid 2000's a number of studies have been released studying the influence of the mycobiome on health and disease, linking presence of certain fungi as well as the relative abundance of fungi in the microbiota to diseases such as cancer and IBD (Aykut et al., 2019; Lewis et al., 2015; Richard and Sokol, 2019). As an example, species belonging to the genus *Malassezia* have been shown to promote pancreatic cancer, and in patients with Crohn's disease the relative fungal abundance has been shown to be increased in expense of the abundance of bacteria (Aykut et al., 2019; Sokol et al., 2017). Interestingly, there seems to be little consensus between results of studies on the mycobiome composition in healthy patients, making it impossible to draw conclusions on the fungal contributors to a healthy microbiota to this day (Sam et al., 2017). To reach such a conclusion there is a need of large-scale studies, with large amounts of data, as the current studies on the mycobiome are often limited to smaller sample sizes. Utilizing the already available metagenomic data from previous large-scale studies focusing on the bacterial microbiome might give a further understanding of the involvement of fungi in health and disease.

2.4.1 Fungal metagenomics

As lack of data is a major hindrance of further exploration of the mycobiome, some studies have focused on alternatives of 18S sequencing for quantification of fungi in the gut, namely through analysis of whole metagenome sequencing data (MGS) (Donovan et al., 2018; Lewis et al., 2015). Historically, most studies of the bacterial microbiota apply 16S sequencing, amplifying a rRNA sequence specific to different bacterial species, to quantify the bacterial microbiota (Tang, 2019). 18S sequencing relies on the same methodology but is instead specific to fungi. Both 16S and 18S sequencing methods are affordable and work well for quantification of their respective target, but these approaches are limited in the sense that the data produced only encompasses one clade in the tree of life, e.g. bacteria or fungi and that there is possible bias depending on the primers used for amplification (Jovel et al., 2016). Fecal MGS is the process of sequencing all the genetic material present in a fecal sample, allowing for unbiased capture of DNA (Jovel et al., 2016). MGS has gradually become more affordable as sequencing prices have plummeted over the last few years, making it increasingly used for studying the bacterial gut microbiome (Jovel et al., 2016; Tang, 2019). By mapping metagenomic sequencing reads to genomes of already sequenced organisms it is possible to identify species present, as well as their abundance by looking at the relative number of reads mapping to the respective genomes (Jovel et al., 2016; Tang, 2019). MGS data can also be used to assemble genomes *de novo*, using the number of assembled contigs to quantify abundance of organisms while allowing for discovery of new species (Jovel et al., 2016; Tang, 2019). MGS data, although often collected for the purpose of studying bacterial composition, can therefore also be applied to analyze fungal, viral, archaeal, and protozoan composition, owing to their unbiased nature (Jovel et al., 2016). Although unbiased, the resolution of non-bacterial sequences is low in MGS data due to bacterial reads making up ~ 99.1% of all reads (Qin et al., 2010; Richard and Sokol, 2019). In the end, fungal sequencing reads have been shown to on average make up less than 0.1% of the total DNA, making it hard to identify anything but the most abundant fungal species (Qin et al., 2010; Richard and Sokol, 2019). This low abundance also hampers *de novo* assembly approaches, as there simply are too few reads for properly assembling fungal chromosomes. Additionally, it is important to keep in mind that DNA extraction protocols used to collect DNA for metagenomic sequencing are usually less harsh protocols adapted for bacteria, making it uncertain if all fungal genomes are extracted before sequencing. Even so, this could in theory be countered by sequencing deeper, as this would produce an increase in the total amount of fungal reads.

The integrative Human Microbiome Project (iHMP) program by the National Institutes of Health (NIH) collected fecal metagenomic data from 127 IBD patients and heathy subjects at several timepoints during the course of a year (Lloyd-Price et al., 2019). The resulting dataset, containing fecal metagenomic data from 1338 samples, is longitudinal and enables analysis of patient specific microbiota compositions as well as changes in the microbiota over time in healthy subjects respectively IBD patients (Lloyd-Price et al., 2019). Herein, an already existing pipeline for identification of fungi in metagenomic data will be modified to make both the analysis and space usage more efficient, after which the IBD data from iHMP will be analyzed. Although fungal metagenomic studies have been conducted previously for Crohn's disease and the mycobiome composition has been quantified through 16S sequencing in IBD patients, this is to our knowledge the largest number of samples in a metagenomic study of the mycobiome to this date. Using results from previous studies as a reference, we hope to establish functionality of the pipeline as well as possible new discoveries.

3 Methods

3.1 Genetic engineering of microbial organisms

3.1.1 Strains and media

The *S. boulardii* strain used in this study was ATCC® MYA-796™, a diploid and prototrophic wild-type strain, as well as strain derivatives in Table XX created in this study. All cultivations were done at 37 °C. Liquid cultivation was done in standard complex yeast-peptone-dextrose (YPD) medium (10 g/L yeast extract, 20 g/L peptone from meat, 20 g/L glucose) or minimal (Delft) medium. Delft medium was composed of (L^{-1}) 22 g glucose, 14.4 g KH_2PO_4, 7.5 g $(NH_4)_2SO_4$, 0.5 g $MgSO_4 \cdot 7H_2O$, 1 mL vitamin solution (50 mg/L D-biotin, 1g/L thiamin-HCl, 1g/L pyridoxin-HCl, 1 g/L D-pantothenic acid hemicalcium salt, 1g/L nicotinic acid, 200 mg/L 4-aminobenzoic acid, 25 g/L m-inositol, pH = 6.5), 2 mL trace metal solution (3 g/L $FeSO_4 \cdot 7H_2O$, 4.5 g/L $ZnSO_4 \cdot 7H_2O$, 4.5 g/L $CaCl_2 \cdot 2H_2O$, 1 g/L $MnCl_2 \cdot 4H_2O$, 300 mg/L $CoCl_2 \cdot 6H_2O$, 300 mg/L $CuSO_4 \cdot 5H_2O$, 400 mg/L $Na_2MoO_4 \cdot 2H_2O$, 1 g/L H_3BO_3, 100 mg/L KI, 19 g/L $Na_2EDTA \cdot 2H_20$, pH = 4) and pH was adjusted to 6 with KOH prior to autoclavation. Vitamin solution and trace metal solution were added to media after autoclavation.
Screening for strains with uracil auxotrophy was done on YPD plates with added 5-fluoroorotic acid (5-FOA), screening for strains carrying plasmids with Kanamycin (KanMX) marker was done on YPD plates with 200 µg/mL added geneticin (G418), and plasmid curation of KanMX marker plasmids was done on YPD plates. All YPD plates had 20 g/L agar. Selection for strains carrying plasmids with auxotrophic markers as well as counterselection-based screening for histidine auxotrophs was done on synthetic complete media (SC) plates, (0.77 g/L Complete Supplement Mixture Drop-out for His respectively Ura from Formedium, 6.9 g/L Yeast Nitrogen Base without amino acids, 20 g/L D(+)-glucose, pH adjusted to 5.5-6.0 with 5 mM NaOH).

E. coli strain DH5α was used in all standard cloning, cultivated in Lysogeny Broth (LB; 10 g/L peptone from casein, 10 g/L NaCl, 5 g/L yeast extract) at 37 °C, with 100 µg/mL ampicillin added for plasmid selection when needed. LB plates with added ampicillin for plasmid selection had the same media composition, with 16 g/L added agar.

Table 1. All strains used in this project are listed below, as well as description of strain genotypes, eventual plasmids carried by strains, and origin of strain.

Strains	Genotype	Plasmids	References
MYA-796	*Wild type*	-	(McCullough et al., 1998)
MYA.u3	MYA-796; *ura3Δ*	-	This study
MYA.a9	MYA-796; *aro9Δ*	-	This study
MYA.aT	MYA-796; P_{AGP1}*Δ::*P_{TEF1}	-	This study
MYA.aT.a9	MYA-796; P_{AGP1}*Δ::*P_{TEF1} *aro9-*	-	This study
MYA.a9PAL	MYA-796; *aro9Δ::*P_{TEF1}*-PAL-*T_{CYC1}	-	This study
MYA.u3.a8	MYA.u3; *aro8Δ*	-	This study
MYA.u3 + pTEF	MYA.u3	p416TEF	This study
MYA.u3 + pTEF_GFP	MYA.u3	p416TEF-GFP	This study
MYA.u3 + pTEF_PAL	MYA.u3	p416TEF-PAL	This study
MYA.u3 + pTEF_PAH	MYA.u3	p416TEF-PAH	This study
ALE5	MYA.u3; 5 days ALE	p416TEF-GFP	This study
ALE16	MYA.u3; 16 days ALE	p416TEF-GFP	This study
ALE34.1	MYA.u3; 34 days ALE	p416TEF-GFP	This study
ALE38.1	MYA.u3; 38 days ALE	p416TEF-GFP	This study

3.1.2 Construction of plasmids and donor fragments

All plasmids used in this study are listed in Table 2 below, and primers are found in Table S1 in the Supplementary materials. All gBlocks and primers were ordered from Eurofins Genomics. The first CRISPR-Cas9 plasmid for genomic editing with *URA3* gRNA was constructed using pCas9-KanMX amplified with primers F-HR/R-HR as backbone, with P_{SNR2}-gRNA-gRNAscaffold-T_{SUP4} flanked by homologous regions PCR amplified using primers fw_gRNA_URA3/rv_gRNA_URA3 from a gBlock. The plasmid was assembled through Gibson assembly using 100 ng plasmid backbone and 300 ng insert incubated at 50 °C during 50 min followed by *E. coli* transformation with 2 µL reaction product. To avoid false positive colonies the plasmid backbone used as template was digested prior do assembly through addition of 1 µL DpnI and 40 min incubation at 37 °C. This plasmid was used as template for all other pCas9-KanMX plasmids, as it includes the gRNA promoter, scaffold and terminator. Other gRNA plasmids were constructed through polymerase chain reaction (PCR) of the pCas9-KanMX respectively pECAS9-KIURA plasmid backbone, using a forward primer binding to the gRNA scaffold with a tail containing the gRNA for the desired gene location followed by an overlap to the reverse primer, as well as a reverse primer binding to the *SNR2* promoter. See primer pairs fw_gRNAPlas_(target gRNA)/rv_gRNAPlas, where the rv_gRNAPlas was kept constant. The linear plasmid fragment was assembled through *E. coli* transformation with 2 µL reaction product. To avoid false positive colonies the plasmid backbone used as template was digested prior do assembly through addition of 1 µL DpnI and 40 min incubation at 37 °C. Confirmation of plasmids was achieved by sequencing of plasmids extracted from transformants.

Plasmids were also constructed for plasmid-based expression of certain genes and for assembly of repair fragments for later use as PCR templates. These include all the p416TEF based plasmids constructed in this study. All genes were expressed under the *TEF1* promoter and *CYC1* terminator. Heterologous genes *PAL* (*Rhodosporidium toruloides*) and *PAH* (*H. sapiens*) were codon optimized and ordered as gBlock. Gene blocks were amplified with tailed primers,

fw_PAL-bindTEF/ rv_PAL-bindCYC respectively fw_PAH-bindTEF/rv_PAH-bindCYC, with tails overlapping P_{TEF1} and T_{CYC1} for construction of plasmids p416TEF-PAL and p416TEF-PAH. PCR products were purified using gel extraction. The plasmid backbone was digested overnight with XhoI and XbaI, gel purified, and the plasmid was assembled using Gibson assembly with 100 ng plasmid backbone and 300 ng insert incubated at 50 °C during 50 min. Construction of p416TEF-PALPAH was done using p416TEF-PAL as a backbone and inserting P_{TEF1}- *PAH* -T_{CYC1} amplified from p416TEF_PAH. The backbone was amplified using primer pair fw_p416_PALPAH/ rv_p416_PALPAH and the insert using fw_CYC_PALPAH/rv_TEF_PALPAH. Confirmation of plasmids was achieved by initial screening by plasmid digestion of plasmids extracted from transformants, followed by sequencing of positive plasmids.

Table 2. All plasmids used in this project are listed below, with descriptions of plasmid backbones and inserts, plasmid origin, and status of construction at the end of the project.

Plasmid	**Backbone; Description**	**References**	**Construction Status**
pCfB2312	CEN/ARS, P_{TEF1}-*Cas9*, *KanMX* marker	(Stovicek et al., 2015)	-
pCas9_KanMX_gHIS	pCfB2312; P_{SNR2}-*HIS3*_gRNA-T_{SUP4}	This study	Complete
pCas9_KanMX_gURA3	pCfB2312; P_{SNR2}-*URA3*_gRNA-T_{SUP4}	This study	Complete
pCas9_KanMX_gARO8	pCfB2312; P_{SNR2}-*ARO8_gRNA*-T_{SUP4}	This study	Complete
pCas9_KanMX_gARO9	pCfB2312; P_{SNR2}-*ARO9_gRNA*-T_{SUP4}	This study	Complete
pCas9_KanMX_gAGP1p	pCfB2312; P_{SNR2}-P_{AGP1}_gRNA-T_{SUP4}	This study	Complete
pCas9_KanMX_gPHA2	pCfB2312; P_{SNR2}-*PHA2*_gRNA	This study	Complete
pECAS9-KIURA	2µ ori, P_{TEF1}-*ECas9* , *KIURA* marker, P_{SNR2} gRNA scaffold	SysBio	-
pECAS9-KIURA_gARO8	pECAS9-KIURA; ARO8_gRNA	This study	Complete
p416TEF	CEN6, *TEF1* promoter, *URA3* marker	SysBio	-
p416TEF-GFP	p416TEF; P_{TEF1}-*GFP*	SysBio	-
p416TEF-PAL	p416TEF; P_{TEF1}-*PAL*	This study	Complete
p416TEF-PAH	p416TEF; P_{TEF1}-*PAH*	This study	Complete
p416TEF-PALPAH	p416TEF; T_{CYC1}-*PAL*-P_{TEF1}-P_{TEF1}-*PAH*-T_{CYC1}	This study	Incomplete

Plasmids were extracted and purified using the GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific). Genes were amplified through Phusion polymerase PCR, or PrimeStar polymerase PCR. Colony PCR was performed using Taq polymerase. PCR products that were to be used in further cloning or to be sequenced were purified using the GeneJet PCR Purification Kit (Thermo Fisher Scientific) or gel extracted when necessary using the GeneJet Gel Extraction Kit (Thermo Fisher Scientific).

Donor fragments for Cas9-mediated knock-in of genes in genomic locations were amplified from their respective plasmid template with primers that had a 60 bp tail overlapping the regions flanking the target locus. The donor fragment for knock-in of P_{TEF1} in the P_{AGP1} location was amplified from p416TEF using primer pair fw_TEF1_AGP1don/rv_CYC1_AGP1don. The donor fragment for insertion of P_{TEF1}-*PAL*-T_{CYC1} in the genetic locus of *ARO9* was amplified from p416TEF_PAL using primer pair fw_TEF-ARO9don/rv_CYC-ARO9don. The donor fragment for insertion of P_{TEF1}-*PAH*-T_{CYC1} into the genetic locus of *ARO8* was amplified from p416TEF_PAH using primer pair fw_TEF-ARO8don/rv_CYC-ARO8don, although this fragment was never successfully amplified. All PCR amplified fragments were PCR purified before transformation.
The forward and reverse strand of repair fragments for gene knockouts were ordered as single stranded 120 bp fragments, overlapping 60 bp on each end with the respective sides of the genomic target to be deleted. These single stranded fragments were assembled through adding 1 µL of each oligonucleotide to 98 µL of MilliQ water, boiling the mixture at 96 °C for 30 min, and letting it cool down to room temperature.

3.1.3 *E. coli* transformation

Chemically competent *E. coli* DH5*α* from the -80 °C freezer where thawed on ice approximately 20 min, after which DNA was added to tubes (2 µL Gibson mix/1µL circular plasmid/ 2µL linearized plasmid), contents were mixed by gentle flicking, and cells incubated 30 min on ice. Cells were then heat-shocked 30 s in a 42 °C water bath, incubated on ice another 2 min, after which 1 mL LB medium was added to tubes. Cells were incubated 1 h for recovery, followed by plating of 100 µL tube contents on a preheated LB+Amp plate and incubation at 37 °C. For plasmid extraction, individual colonies were picked and inoculated in LB+Amp media and grown overnight before plasmids were extracted using the GeneJet Plasmid Miniprep Kit (Thermo Fisher Scientific). Correctly assembled plasmids were screened for using restriction digestion, followed by sequencing. The exception were the gRNA plasmids that were sequenced right away due to there being no appropriate restriction sites.

3.1.4 Establishing transformation protocols for *S. boulardii*

Transformation of yeast was tested both using variants of the LiAc/SS carrier DNA/PEG method and using electroporation. The LiAc/SS carrier DNA/PEG method was based on the paper published by Gietz and Schiestl in 2007, with a 30-minute incubation step added before heat shock (Gietz and Schiestl, 2007). In order to optimize efficiency of transformation for *S. boulardii* the incubation and heat shock time, recovery time, and the amount of DNA added were modified throughout the project. Incubation time was varied between 30 min and 1 h, heat shock between 20 min and 1 h, recovery time between 0 h and overnight recovery, and DNA added for Cas9-mediated genome edits were 0.5-2 µg plasmid and 1-4 µg insert. As optimization of protocols were not the main aim of the thesis, no large-scale evaluation experiments were conducted. Modifications of the protocol were only tested in transformations included in the main project. In the final protocol used throughout the end of the thesis, incubation time was 30 min followed by 45 min heat shock, the recovery was extended to overnight, and the amount of DNA added was 2 µg plasmid and 4 µg repair fragment. For transformation of plasmids with auxotrophic marker, 1 µg plasmid was used. Furthermore, on recommendation from Alex Hedin who is working with the same organism in a different lab, 30 °C was used for incubation and recovery instead of 37 °C as was initially used in this project.

Electroporation was tested as an alternative to the LiAc/SS carrier DNA/PEG method with the hopes of reaching a higher efficiency (Thompson et al., 1998). The amount of DNA added was 2 µg plasmid and 4 µg repair fragment.

3.1.5 Transformant screening

In order to screen for transformants with successful genomic modifications colony PCR and re-streaking on selective plates was applied. In screening for histidine- and uracil-auxotrophic strains, colonies were re-streaked first on SC dropout medium and then on YPD.
In screening for other genomic modifications PCR was applied. Genomes were extracted by adding 15 µL 20 mM NaOH to tubes from the previous step, followed by boiling at 96 °C for 30 min, 2x30 s vortex, and centrifugation to spin down cell debris. 2 µL of the supernatant was used in colony PCR, for which all primers are listed in Table S1 in Supplementary materials with colPCR and the name of the target genomic location in the primer name.

3.1.6 Confirming strains as *S. boulardii*

As *S. boulardii* is physiologically similar to *S. cerervisiae* and both are cultivated in the lab, it is important to be able to distinguish these if a contamination is suspected. In this study both the ITS1-ITS4 region of the fungal genome and two marker genes were used for this purpose. The ITS region is shorter in *S. boulardii* compared to *S. cerevisiae* and comparing several strains of the respective yeasts it was found that *PDR18* is unique to *S. boulardii* and *PGA3* is unique to *S. cerevisiae.* Primer pair ITS1/ITS4 was used for amplification of the ITS region, primer pair fw_SbMARK_PDR18/rv_SbMARK_PDR18 for amplification of *PDR18* and fw_ScMARK_PGA3/rv_ScMARK_PGA3 for amplification of *PGA3.*

3.2 Adaptive laboratory evolution

3.2.1 Strains, media, and growth conditions

Two separate ALE experiments were performed to increase phenylalanine consumption in *MYA.u3-GFP*, using different sets of modified synthetic minimal Delft media. In the first ALE, a series of media with decreasing glucose levels (1%, 0.8%, 0.6%, 0.4%, 0.2%, 0.1%, 0%), a constant level of 1% phenylalanine, and without nitrogen was used. The second ALE utilized the same media composition overall, but with a constant level of 0.6% glucose. In this ALE, the aim was for all glucose to be consumed within 24 h but not less, which could be achieved with a level of 0.6 % based on growth in previous ALE. Furthermore, the medium in the second ALE was filtered instead of autoclaved.
The medium was composed of (L^{-1}) (11, 8.8, 6.6, 0.4, 0.2, 0.1, 0) g glucose, 11 g L-phenylalanine, 14.4 g/L KH_2PO_4, 0.5 g $MgSO_4 \cdot 7H_2O$, 1 mL vitamin solution (50 mg/L D-biotin, 1g/L thiamin-HCl, 1g/L pyridoxin-HCl, 1 g/L D-pantothenic acid hemicalcium salt, 1g/L nicotinic acid, 200 mg/L 4-aminobenzoic acid, 25 g/L m-inositol, pH = 6.5), 2 mL trace metal solution (3 g/L $FeSO_4 \cdot 7H_2O$, 4.5 g/L $ZnSO_4 \cdot 7H_2O$, 4.5 g/L $CaCl_2 \cdot 2H_2O$, 1 g/L $MnCl_2 \cdot 4H_2O$, 300 mg/L $CoCl_2 \cdot 6H_2O$, 300 mg/L $CuSO_4 \cdot 5H_2O$, 400 mg/L $Na_2MoO_4 \cdot 2H_2O$, 1 g/L H_3BO_3, 100 mg/L KI, 19 g/L $Na_2EDTA \cdot 2H_20$, pH = 4) and pH was adjusted to 6 with KOH prior to autoclavation. Vitamin solution and trace metal solution were added to media after autoclavation.
Growth conditions were the same for both ALE experiments, where cultures of 10 mL medium in 100 mL shake flasks where incubated in 37 °C at 200 rpm. In the first ALE, one overnight culture was inoculated in regular Delft medium, washed and resuspended in MilliQ water after 16 h, and diluted in 5 parallel shake flasks to an OD_{260} of 0.01 in 10 mL 1% glucose 1% phenylalanine media. The following 34 days the cultures were diluted to an OD_{260} of 0.5 every

24 h, lowering the glucose content of the media every time the cell concentration had stabilized and began to increase between dilutions. The last 6 days, cells were diluted every 48 h instead due to low growth in the 0.1% glucose media. Every 7 days, cultures were controlled for contamination by microscopy and a cryostock was prepared for each of the 5 parallel cultures.
The second ALE was inoculated from cryostocks from day 16 of the first ALE in order to save time, as cells had already been cultivated in media containing 0.6% glucose several days at point of crystock collection. For each of the 5 ALE strains, about 250 µL of cryostocks were inoculated in 10 mL Delft media in order to conserve biodiversity from the previous ALE. After 24 h cells were spun down at 4200 rpm, washed in 25 mL sterile MilliQ water, spun down and eluted in 1 mL sterile MilliQ. OD was measured and volume was adjusted so that all cell suspensions had an OD of 0.5. 1 mL of each cell suspension was then inoculated in 9 mL of ALE media. Instead of diluting to a fixed OD every 24 h, a fixed volume of 1 mL was used instead. The OD was still recorded at each sampling, cultures were checked for contamination by microscopy every 7 days, and a cryostock was prepared for each culture in parallel with contamination check.

3.2.2 Fluorescence microscopy

As strains used in the laboratory evolution carried the p416TEF_GFP plasmid, cultures were checked for contamination once per week by fluorescence microscopy. If the majority of strains in the population were fluorescent it was assumed that culture was non-contaminated. The microscope used was a Leica DMI4000B, with GFP filter cube Exiter at 437/30, Dichronic at 495, and Emitter at 520/35.

3.2.3 Growth assay

On two different media, Delft supplemented with 100 mg L-phenylalanine and the medium used in the second ALE, growth was evaluated for the strains developed in this study. Growth of MYA.aT, MYA.a9, MYA.aT.a9, MYA.a9PAL strains were compared to growth of MYA-796 WT strain, growth of MYA.u3+p416TEF_PAL and MYA.u3+p416TEF_PAH was compared to growth of MYA.u3+p416TEF. For all these strains except for MYA.a9PAL, three transformation replicates were used, together with three assay replicates for each initial culture. As only one transformation replicates were available for MYA.a9PAL, two strain replicates was used instead.
Growth of the final ALE strains, ALE 34.1 and ALE 38.3, as well as intermediate strains ALE 5 and ALE 16, was compared to growth of MYA.u3+p416TEF_GFP. Cryostocks were streaked on 1% phenylalanine 0.6 % glucose plates, from which colonies were picked for the growth assay. Since 5 parallel cultures were used in the ALE, only one colony per each of the 5 ALE cultures could be evaluated, with three assay replicates per colony.
The growth assay was performed in a Growth Profiler 960 from Enzyscreen on 4 separate 96-well plates, with non-randomized sample order. MYA strains were analyzed on the same plate as the control strain, which was also the case for MYA.u3 strains, and ALE strains. Due to the number of cultures, ALE strains had to be divided over two plates, both containing the control stain as well.
For precultures all strains were inoculated in 5 mL Delft medium and incubated 48 h in 37 °C at 200 rpm. For each culture, the OD was measured and a culture volume corresponding to OD 0.5 in 1 mL MilliQ was transferred to a sterile 1.5 mL Eppendorf tube. Contents were spun down at 4500 rpm, supernatant discarded, and cells were re-eluted in 1 mL MilliQ water. 25 µL of cell suspension was then added to 225 µL medium in designated wells on 96-well plate. Growth of strains was evaluated on both media, over a time course of 70 h.

3.2.4 LC-MS/MS evaluation of phenylalanine consumption

3.2.4.1 Media and strains

Initial evaluation of phenylalanine consumption was done in Delft medium with 50 mg/L respectively 500 mg/L supplemented phenylalanine at timepoints 12 and 24 h for strains MYA.u3+p416TEF, MYA.u3+p416TEF_PAL and MYA.u3+p416TEF_PAH. Based on the results from this experiment it was decided to use Delft media containing 100 mg/L phenylalanine in the final assay.

In the assay samples were prepared in two rounds, the first round for all the non-ALE strains and the second round for all the ALE strains. For strains MYA, MYA.aT, MYA.a9, MYA.aT.a9, MYA.u3+p416TEF, MYA.u3+p416TEF_PAL and MYA.u3+p416TEF_PAH, three transformation replicates were used, together with two assay replicates for each sample.

For MYA.a9PAL, which had only one transformation replicate, two strain replicates and two assay replicates were used instead.

For ALE strains one colony was picked per parallel ALE culture for each of ALE 5 1-5, ALE 16 1-5, ALE 34.1 1-5, and ALE 38.2 1-5. For the control strain MYA.u3+p416TEF_GFP, three transformation replicates were used. Two assay replicas were prepared for each of the collected samples.

3.2.4.2 Sample collection

Samples were collected in three rounds, first for MYA, MYA.aT, MYA.a9, MYA.aT.a9, and MYA.a9PAL, then for MYA.u3+p416TEF, MYA.u3+p416TEF_PAL and MYA.u3+p416TEF_PAH, and lastly for the ALE strains and controls.

Precultures were started by inoculation of stains and strain replicates in 5 mL Delft media and incubating these 48 h in 37 °C at 200 rpm. OD was then measured and a volume of culture corresponding to OD 5 in 1 mL was transferred to a new, sterile Eppendorf tube, pelleted at 4200 rpm 5 min, and resuspended in 1 mL sterile MilliQ. 100 µL cell suspension was then added to 9.9 mL media in a 100 mL shake flask to reach an initial OD of 0.05. Shake flasks were incubated at 37 °C at 200 rpm.

OD was measured and samples collected at 12 and 24 h after inoculation. 1 mL culture was transferred to a 1.5 mL Eppendorf tube, cells pelleted, and supernatant collected in a cryovial. Collected culture supernatant was placed in a -80 °C freezer within 2 h of sample collection.

3.2.4.3 Sample preparation

Sample preparation was done in a randomized sample order, with quality controls every 12 samples. All reagents used were included in the aTRAQ kit. A quality control sample was prepared through transfer of 50 µL of each sample included in the batch into the same tube, and baseline sample for the assay was prepared by transferring 1 mL of the assay medium into a new tube.

Frozen samples were thawed in a fridge overnight and vortexed. 40 µL of each physiological sample, QC and baseline sample were added into a first batch of Eppendorf tubes, two assay replicas of physiological respectively baseline samples were prepared. Randomization of sample order begins at this step. 10 µL of sulfosalicylic acid was added to each tube, tubes were shortly vortexed and then centrifuged at 10 000 g for 2 min. 10 µL of supernatant was transferred to a new batch of Eppendorf tubes to which 40 µL of labelling buffer was added, followed by a short vortex and then centrifugation at 10 000 g for 2 min. 10 µL of the supernatant was transferred to a new batch of tubes, and tubes from the previous step were stored in a refrigerator as a new sample preparation can begin from this intermediate step. Contents of 7 aTRAQ Reagent Δ8 vials were spun down at 10 000 g for 2 min, 70 µL of isopropanol was added, tubes were vortexed, and contents spun down at 10 000 g for 2 min.

5 µL of the prepared aTRAQ Reagent Δ8 solution was added to each of for the sample tubes, samples were shortly vortexed and centrifuged at 100 g for 2 min, followed by incubation at room temperature 30 min. Hydroxylamine tubes from the kit was shortly vortexed and contents spun down, before 5 µL was added to each sample. Each sample was shortly vortexed and spun down at 10 000 g for 2 min, followed by incubation at room temperature for 15 min.
Internal standard was prepared by adding 1 mL of standard diluent to aTRAQ internal standard vial, which was vortexed 30-60 s until contents were dissolved, after which another 0.8 of standard diluent was added. Tubes were vortexed again in order to mix the contents.
32 µL of the prepared internal standard solution was added to each sample tube, followed by 400 µL MilliQ water. Samples were vortexed and centrifuged at 10 000 g for 2 min, after which contents were transferred to autosampler vials.
100 µL of prepared internal standard solution was added to a separate vial, to be used in initial testing of the system for adjustment of retention times.

3.2.4.4 Method and instrumental settings

A LC-MS/MS based method for measurement of phenylalanine in serum and plasma samples was utilized for measurement of phenylalanine consumption, using the aTRAQ™ Kit for Amino Acid Analysis of Physiological Fluids from Sciex. The system used was Shimadzu Nexera LC with an 18C column with dimensions 4.5x150 mm and a Sciex Qtrap 6500+. The LC system was set up according to the aTRAQ kit manual, with a column temperature of 50 °C, a flow rate of 0.8mL/min, and a LC gradient as described in Table 3.

Table 3. LC settings to be used in the aTRAQ method, describing the percentage of the total flow constituted by each of the Mobile phases at the different time points.

Total time (min)	% Mobile phase A	% Mobile phase B
0	98	2
6	60	40
10	60	40
11	10	90
12	10	90
13	98	2
18	98	2

Table 4. MS settings for identification of phenylalanine respectively internal standard for quantification of phenylalanine.

ID	Q1 Mass (Da)	Q3 Mass (Da)	Retention time	Collision energy (volts)
Phe	314.2	121.1	6.9	30.0
Phe internal standard	306.2	113.1	6.9	30.0

Mobile phase A for 75 samples was prepared through adding 1 mL aTRAQ mobile phase modifier A and 0.5 mL mobile phase modifier B to 500 mL MilliQ water, mixing contents, and adding another 500 mL MilliQ water before transferring contents to a flask. Mobile phase B for 75 injections was prepared through adding 0.5 mL aTRAQ mobile phase modifier A and 0.25 mL mobile phase modifier B to 250 mL methanol, mixing contents, and adding 250 mL methanol before transferring contents to a flask. Rinsing solution was prepared by adding 3

parts of methanol, 3 parts acetonitrile, 1 part isopropanol, 1 part milliQ water, and 0.01 part formic acid into a 1 L flask.
An autosampler vial with 1 mL Mobile Phase A was prepared to be used as blank throughout the assay. System suitability was evaluated through an initial run of 3 internal standard solution samples, and retention times were modified based on results.

3.3 Quantitative bioinformatic analysis of the mycobiome

FindFungi is a pipeline for analysis of the mycobiome in metagenomic data developed by Paus Donovan and colleagues in 2015 (Donovan et al., 2018). The pipeline relies on three gates for identification and confirmation of fungal reads in shotgun metagenomic samples (**Fig. 4**). After initial quality filtering, reads are fed to Kraken2 for alignment to 10 custom fungal databases, built from all representative fungal genomes in NCBI. Kraken2 filters out fungal reads and assigns each read with the predicted taxid. Reads assigned to each taxid are mapped to a taxid specific BLAST database constructed from 10 pseudo-chromosomes of the fungal genome it represents. This allows for quantification of the fraction of pseudo-chromosomes to which reads mapped, along with the calculation of Pearson skewness based on the distribution of mapped reads. These scores are then applied as a final filtering step in the pipeline to avoid false positives in the data. Manual curation of the final dataset may still be necessary however, due to the possibility of badly curated genomes in the Kraken and BLAST databases. These badly curated genomes can, for instance, lead to bacterial reads being misclassified as fungal.

The reason why two different database filtering gates are necessary in FindFungi are the different alignment methods used by the respective databases. While Kraken2 utilizes the faster and less accurate sorted list of k-mer/LCA pairs, BLAST is based on an initial heuristic approach where all genomic locations aligning to the first part of the query sequence are collected, followed by local alignment in each location. Although this method is superior to Kraken2 in terms of accuracy it is slow, and not suitable for filtering of reads in a metagenomic sample. By using both databases fast filtering of fungal reads is combined with a high accuracy of alignment.

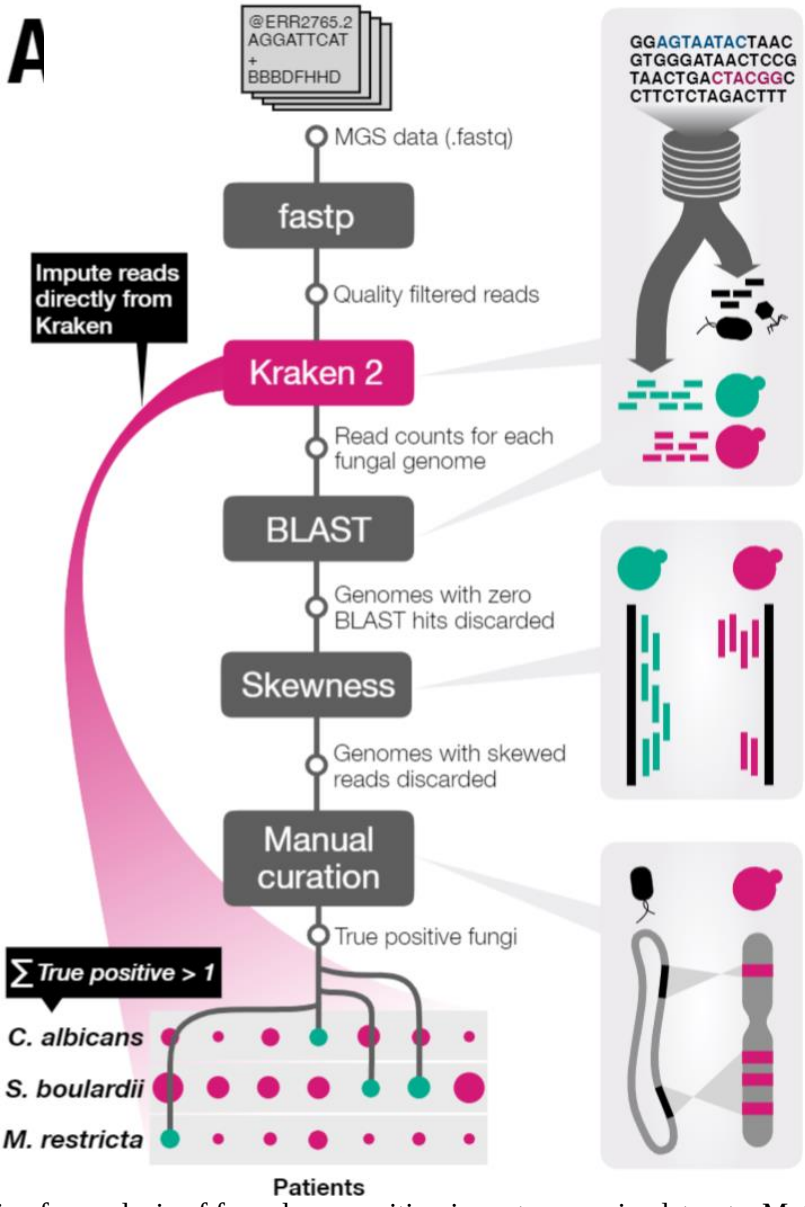

Figure 4. Pipeline for analysis of fungal composition in metagenomic. datasets. Metagenomic shotgun data (MGS) fastq files input to FindFungi pass several filtering gates before manual curation. Reads are quality filtered, passed to the Kraken2 fungal database where fungal reads are separated from the rest of the data and assigned a taxid. All reads predicted to belong to a certain taxid by Kraken2 are passed to BLAST and aligned to the corresponding genome. If reads align, a skewness score is assigned depending on genome coverage and distribution of the mapping reads by which the reads are filtered, followed by manual curation of the data.

3.3.1 Modification of the FindFungi pipeline

The Kraken databases used by FindFungi were updated to Kraken2, a newer version of Kraken. In order to do so the existing script for the database setup had to be modified and was made more user friendly in order to make future updates simple. All representative fungal genomes available in NCBI were included in the new database, expanding the total fungal species in the data from 949 to 1115. As Kraken2 is more space efficient compared to Kraken, the number of databases were decreased to 10, each of 49 GB. The FindFungi scripts were modified to accommodate for this change. The BLAST databases are yet to be updated however, meaning that none of the newly included fungi can be recognized as true positives and will hence not be included in further analysis.

Minor modifications were made of the FindFungi pipeline, including decreasing the run time per sample through submitting all BLAST alignments to one node instead of having one node per alignment, merging subscripts to the main script, an. The output was also modified to not only include the identified fungi, but also the total read counts, total number of fungal reads and sample identity, in order to allow easy concatenation of results from different samples to simplify downstream analysis of the output data.

3.3.2 Metagenomic analysis of the mycobiome

Initially only true positive fungi were included in the data analysis, but as the fungal species predicted as true positives were sparse between samples this resulted in an abundance matrix with mostly zeroes, making analysis difficult. This can in part be attributed to the harsh filtering conditions by the BLAST and skewness steps, both of which become problematic at low fungal read counts. To circumvent this problem, reads from fungi predicted as true positive in at least one sample were included for all samples, even if these were predicted as false positives based on fraction of chromosome mapping and Pearson's skewness score. Fungal read counts were normalized by the total number of sequenced reads in each sample and log2 transformed in order to conform each sample to normality.

Shifts in fungal abundance were investigated through the use of Multi-Dimensional Scaling (MDS) with Euclidean distances and through hierarchical clustering of fungal abundances in samples by Pearson correlation. Shifts in abundance of each individual fungus in the different conditions was examined though violin plots, using an unadjusted two-sided Kruskal-Wallis test to determine significance.

4 Results

4.1 Optimization of genetic engineering of *S. boulardii* and resulting strains

S. boulardii was introduced in the lab for the first time during the course of this project. Prior to the project it was assumed that the similarity between *S. boulardii* and *S. cerevisiae* would allow for efficient application of protocols developed for engineering of *S. cerevisiae* in *S. boulardii* as well. Although engineering using the same protocols was overall successful, the efficiency of Cas9-mediated genome engineering turned out to be lower when applied in *S. boulardii*. The average number of transformants resulting from transformation of *S. boulardii* with pCfB2312-based plasmids were only 4.66 ± 4.67 when using the LiAc/SS carrier DNA/PEG method with recommendations for *S. cerevisiae* (**Tab. 3**). Furthermore, out of a total of 41 screened transformants for *HIS3* and *URA3* knock-out, none had a correct deletion. One of the strains screened for *URA3* knock-out did however have a partial deletion of the gene, causing auxotrophy (**Fig. 4 A, B**). The partial deletion did not include the gRNA target, making it possible that the strain was a result of counterselection on 5-FOA rather than of Cas9-mediated editing (**Fig. 4 B**). However, it is not possible to confirm this. The strain was called MYA.u3.

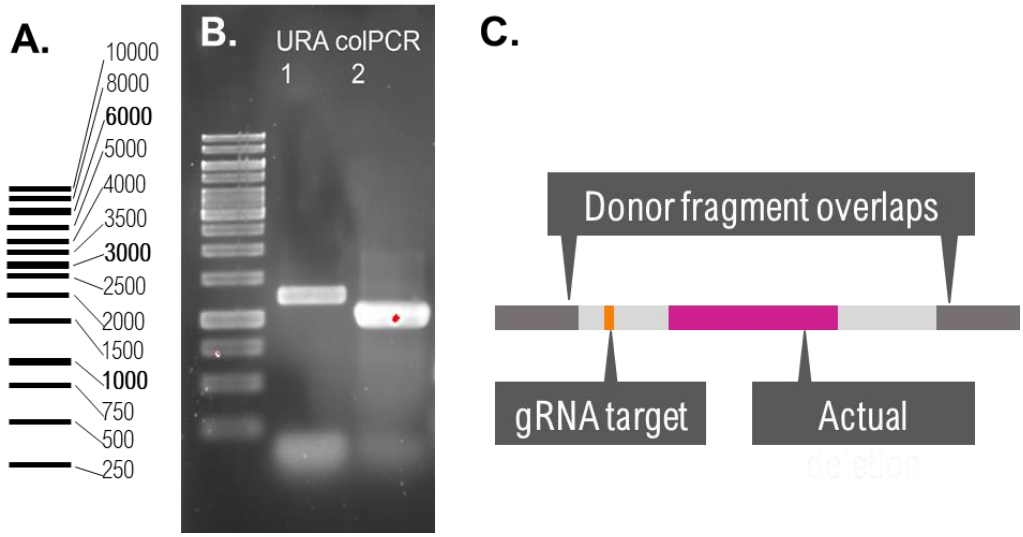

Figure 4. *URA3* deletion in *S. boulardii* (**A**) Length of bands in ladder. (**B**) Length of DNA fragments from colony PCR of *URA3* in two transformants after attempted deletion of the gene. The length of the fragment in lane 1, 1240 bp, corresponds to unmodified *URA3*. If deletion was successful, the band should have been 436 bp. The band in lane 2 does not correspond to any of these lengths, as it is closer to 1000 bp. (**C**) Overview of sequencing results of colony PCR fragment from lane 2 in B. Although the full length of DNA between the donor fragment overlaps should have been deleted, sequencing showed that only part of the sequence had been removed (pink). The deletion did not include the gRNA target (orange).

With the total efficiency being consistently low with varied incubation time, heat shock time, and recovery a new approach was attempted, namely overnight recovery. This resulted in an increase in the number of transformants compared to when a shorter (0-5 h) recovery time was applied (**Tab. 3**). The average number of transformants more than doubled, from 4.56 ± 4.67 to 11.37 ± 12.66, although the variance increased as well due to some transformations still resulting in zero transformants while the highest number of transformants from a transformation were 46 (**Tab. 3**). Using overnight recovery, the efficiency of Cas9-mediated genomic modifications increased as well. While no successful genomic modifications were achieved previously, *ARO9* deletion, *PAL* knock-in in the *ARO9* genomic locus, and P_{TEF1} knock-in in the P_{AGP1} genomic locus were achieved using overnight recovery (**Fig. 5 B, C, D**). The efficiency of modification varied with the gRNA used, as *ARO9* knock-out was successful in 88.2% of screened transformants, *PAL* knock-in in the *ARO9* genomic locus was successful in 2.2% of screened transformants, and P_{TEF1} knock-in in the P_{AGP1} genomic locus was successful in 14.3 % of screened transformants (**Tab. 4**). Even with overnight recovery none of the other attempted genomic modifications were successful, including *ARO8* knock-out, *PHA2* knock-out, and *HIS3* knock-out for which multiple transformations were attempted and a total of 32-39 colonies were screened per strain (**Tab. 4**). *ARO9* knock-out strains were named MYA.a9, strains with *PAL* knock-in in the *ARO9* genomic locus were named MYA.a9PAL, and strains with and P_{TEF1} knock-in in the P_{AGP1} genomic locus were called MYA.aT. *ARO9* knock-out was also successfully implemented in the MYA.aT strain, resulting in strains MYA.aT.a9.

Table 3. Comparison of overnight versus non-overnight recovery in strains transformed with the Cas9 plasmid with the KanMX marker for genomic insertion and deletion. The average number of colonies on the plate after transformation, the standard deviation from average, the minimum number of transformants, and the maximum number of transformants are listed in the table to give an overview of relative efficiency. All transformed strains and all gRNAs used in this project were included in calculations. Note that as continuous optimization was attempted, variations occurred in terms of incubation time,

incubation temperature, heat shock time, and the amount of plasmid respectively donor fragment added during transformation. The non-overnight incubation time varied between 0-5 h.

Overningt/non-overnight recovery	Average number transformants	SD	Min. Number transformants	Max. Number transformants
Non-overnight	4.56	4.67	0	11
Overnight	11.37	12.66	0	46

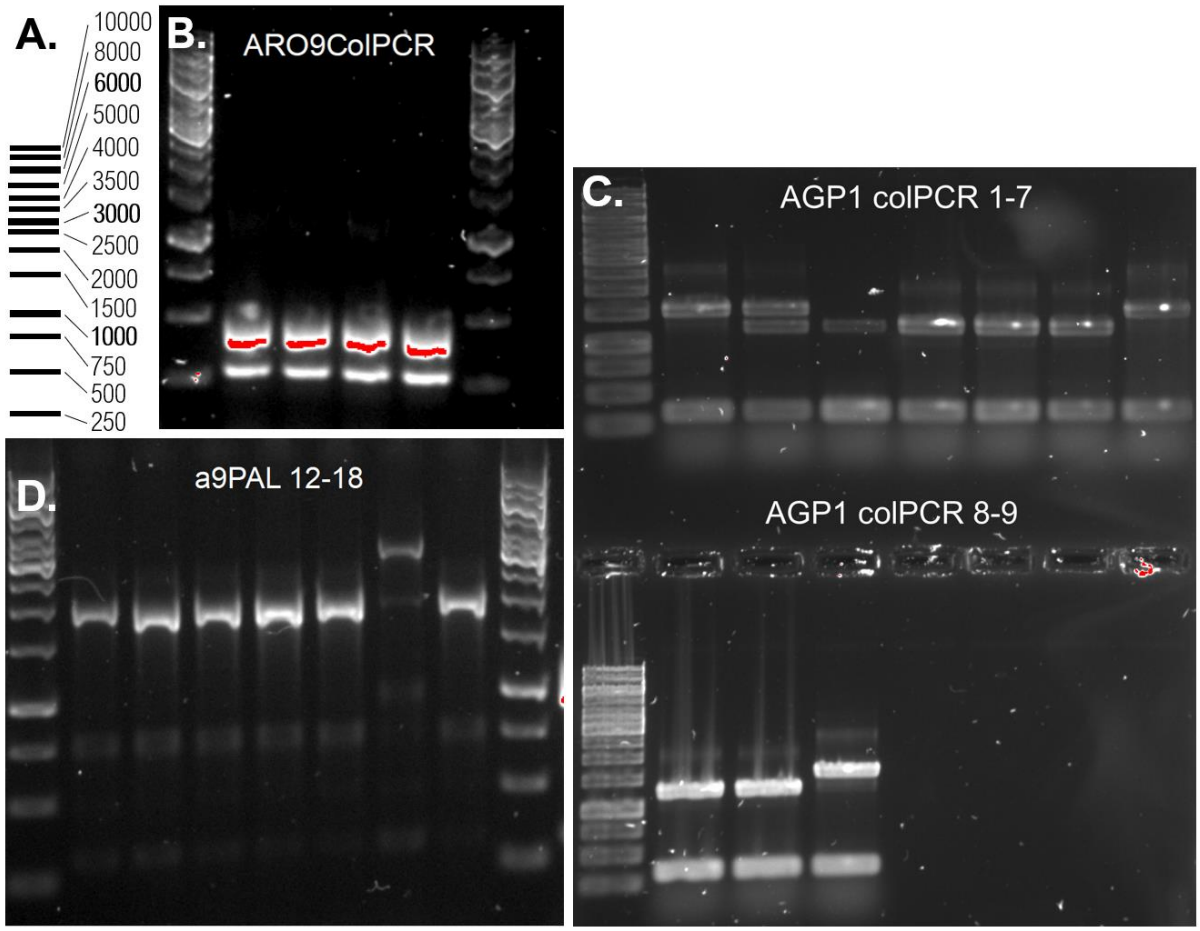

Figure 5. Analysis of *S. boulardii* transformants. **(A)** Length of bands in ladder. **(B)** Gel with the results of colony PCR from 4 colonies screened for *ARO9* deletion. Successful deletion *ARO9* results in a DNA fragment length of 525 bp, while band for the full gene results in a 2050 bp fragment. Although bands appear shorter than 525 bp, subsequent sequencing confirmed the deletion was implemented correctly. **(C)** Gel shows result of P_{TEF1} knock-in in the P_{AGP1} genomic locus. Successful integration results in a 1555 bp fragment while unsuccessful integration results in a 1259 bp fragment. Three of the screened colonies (lane 1, 6, 9) had a correctly implemented genomic modifications, and one of the screened strains (lane 2) had a successful integration in one chromosomal copy while the other retained the native P_{AGP1} shown by the dual bands. A weak band is seen above all bands, of which the origin is unknown. **(D)** Gel with the results of colony PCR of 7 strains screened for integration of *PAL* in the genomic locus of *ARO9*. Successful integration results in a 2600 bp fragment, while the native gene results in a 2050 bp fragment. One colony (lane 6) resulted in a fragment seemingly longer than 2600 bp, but the integration was confirmed by subsequent sequencing. Note however that a weak band is present in the same lane at about 2050 bp, making it possible that the strain retained a genomic copy of *ARO9*. The band is very weak however, making this event unlikely.

Table 4. Overview of efficiency of knock-out respectively knock-in for each gRNA used in this project, when expressed from the Cas9 plasmid with KanMX marker. For each combination of gRNA and donor fragment the total fraction and percentage of screened transformants that were correct are listed. The

fractions are the total number correct transformants divided by the total number of transformants screened during the course of the project. Results were only included from transformations where overnight recovery was applied.

gRNA	Donor	Fraction of correct transformants	% Correct transformants
ARO8	KO don	0/32	0.0
ARO9	KO don	15/17	88.2
ARO9	PAL don	1/46	2.2
HIS3	KO don	0/17	0.0
P_{AGP1}	P_{TEF}	5/35	14.3
PHA2	KO don	0/39	0

Due to the low efficiency of the *ARO8* knockout using the KanMX selection marker, an attempt was made with using *KIURA* as a selection marker towards the end of the project, as auxotrophic markers were recently suggested to result in a higher transformation efficiency compared to antibiotic markers in *S. boulardii* (Durmusoglu et al., 2020). In accordance with these results, the number of transformants as well as the rate of successful of Cas9-mediated knockouts increased using the auxotrophic *KIURA* marker. Using the same gRNA and donor fragment as in previous transformations for *ARO8* knockout, cloning efficiency increased from 0% to over 16.66 % in the first cloning attempt where 1 in 6 screened colonies appeared to have a single chromosome knockout, and 1 in 6 colonies seemed to have a knockout in both genomic copies (**Fig. 6**). Electroporation was also attempted once for *HIS3* knock-out, *ARO8* knock-out, and P_{TEF1} knock-in in the P_{AGP1} genomic locus, but due to zero colonies growing on any of the transformation plates the method was not applied again.

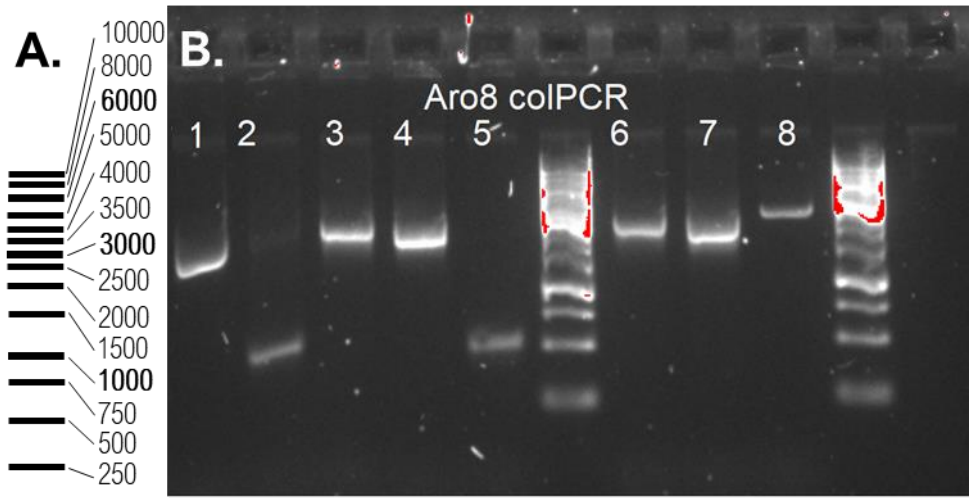

Figure 6. (A) Length of bands in ladder. **(B)** Gel with the results of colony PCR from 8 colonies screened for *ARO8* deletion, attempted with the *KIURA* marker instead of the KanMX marker. Successful deletion *ARO8* results in a DNA fragment length of 530 bp, while presence of the full gene results in a 2038 bp fragment. Two colonies appear to be correct (lane 2 & 5), although a weak band corresponding to full length in lane 2 indicates that this colony might only have a successful deletion in one genomic copy.

Looking at the transformation efficiencies of p416TEF-based expression plasmids into the MYA.u3 strain, with no attempts to genomic modifications, a higher number of transformants were yielded. The number of transformants were around 200-400 per µg plasmid DNA per 10^8

cells in almost all cases (p416TEF, p416TEF-PAL, p416TEF-PAH), except for in transformation of p416TEF_GFP where the number of transformants were 53. Furthermore, overnight recovery was not necessary. Compared to transformation in *S. cerevisiae* this is still a low efficiency however, as the developers of the applied protocol states that 1×10^6 transformants per μg plasmid DNA per 10^8 cells can be expected for most *S. cerevisiae* strains (Gietz and Schiestl, 2007).

4.2 ALE setup and growth comparison of final strains

In order to identify new targets for engineering of *S. boulardii* to become a more efficient Phe consumer, an evolutionary approach was applied. The aim was for the resulting strains to be able to grow solely on Phe, or alternatively to increase the efficiency of Phe uptake. For this purpose, two sets of ALEs were applied using batch evolutions with 5 parallel cultures and dilutions every 24 h (**Fig. 7**).

The conditions used in the first ALE were a Phe concentration set constantly at 1%, a glucose concentration decreasing in steps of 0.2% from 1% to 0%, and no added ammonia in minimal medium. The glucose concentration was decreased as soon as the OD at the current concentration level had stabilized and started to increase, and cultures were diluted every 24 h to OD 0.05. The applied evolutionary pressure in this case was the decreasing availability of the preferred carbon source, glucose, while Phe was present in abundance. OD at each dilution for each strain, as well as the media used at different time points of the ALE, are presented in Supplementary Figure S1**.** Based on the plots in the figure, it is evident that no strain seemed to grow better compared to the others, and although growth improved slightly compared to the initial measurement on each media the improvement was only large on the first medium used, 1% glucose 1% Phe-nitrogen, as well as when strains were grown in 0.6% glucose 1% Phe-nitrogen media. Strains did not appear to adapt well to growth on media with glucose levels below 0.4 % (**Supplementary Fig. S1**).

The second ALE was initiated using part of the cryostock from the 16th d^{ay} of the previous ALE. This was an attempt to save time as the strains from the first ALE at day 16 had been evolved in the same media as applied in the second ALE several days already (**Supplementary Fig. S1**). The ALE strategy was slightly modified, with the medium having a constant glucose level of 0.6 % and a Phe level of 1%, and instead of diluting to the same OD every time, a fixed volume of 1 mL was transferred instead (**Fig. 7**). The strategy was changed in order to allow for more growth compared to in the previous ALE and consequentially also a higher chance of mutations occurring. For this purpose the glucose level was chosen based on the OD from growth in the same media in the previous ALE, ranging between 0.4 and 0.75 (**Supplementary Fig. S1 A**). However, due to the different dilution method applied a larger number of cells were transferred with each dilution, resulting in an OD of over 2 during the larger part of the evolution. Based on the higher OD, there was more glucose available than intended, making it probable that the selective pressure was insufficient during this ALE. This is not certain however, as the initial ODs were higher as well.

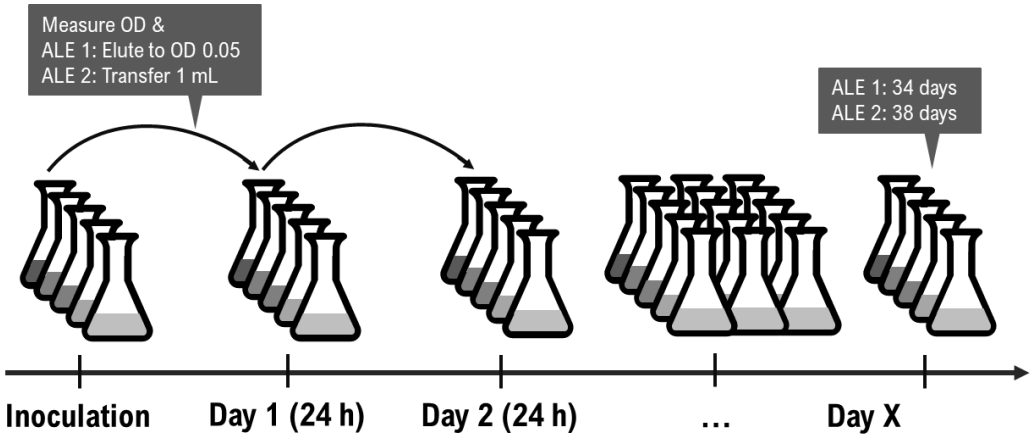

Figure 7. ALE strategy. Two ALEs were performed using different elution techniques. Five parallel flasks were used, and cultures were eluted every 24 hours during 34 respectively 38 days.

The resulting final strains from both ALEs, strain ALE_34.1 from the first ALE and strains ALE_38.2 from the second ALE, as well as intermediate strains ALE_5 and ALE_16 were evaluated in terms of growth to see if any improvement was found. Growth was evaluated on two different media compositions, Delft + 100 mg Phe (2% glucose 0.01% Phe) as well as Delft 0.6% glucose 1% Phe -nitrogen. Comparison of growth on both media should give insight in overall changes in growth on regular media as well as eventual improvement of growth in ALE conditions.

Continuous growth of all strains was recorded, but as the overall differences were small between strains, ODs at timepoints 12, 24, and 36 h were selected for further analysis of significance of differences in growth between strains (**Supplementary Fig. S2 & S3**). These time points were selected as they roughly represent OD at the glucose phase, ethanol phase, and stationary phase. Growth evaluation was performed in a growth profile, and the final and intermediate ALE strains were grown on separate plates together with the control. The results presented in **Figure 8 A, C** respectively **Figure 8 B, D** represent each of the plates. By the difference in growth of the control strain, inoculated from the same precultures, on the different plates it is evident that OD was recorded differently for the different plates. Based on this, results should be compared within plots only.

Examining ODs on both media after 24 h of growth for intermediate strains ALE_5 and ALE_16, no remarkable conclusions could be drawn for both strains across timepoints and media (**Fig. 8 A, C; Supplementary Tab. S2**). The most notable result was the increased growth of strain ALE_16. Although there was a lot of variance in the data from growth of Delft+100 mg Phe and although the increase the increase was small on Delft 0.6% glucose 1% Phe -nitrogen, the increase in OD was significant (p-value < 0.001; **Fig. 8 A; Supplementary Tab. S2**). Looking instead at strains ALE_34.1 and ALE_38.2, we observed increased OD in the Delft+100 mg media at the 24 h mark compared to the control strain (**Fig. 8 B; Supplementary Tab. S3**). This was also the case for growth on Delft 1% Phe + 0.6% glucose – nitrogen media, but here the OD of strain ALE_38.2 was also significantly increased compared to OD of strain ALE_32.1 (**Fig. 8 D; Supplementary Tab. S3**).

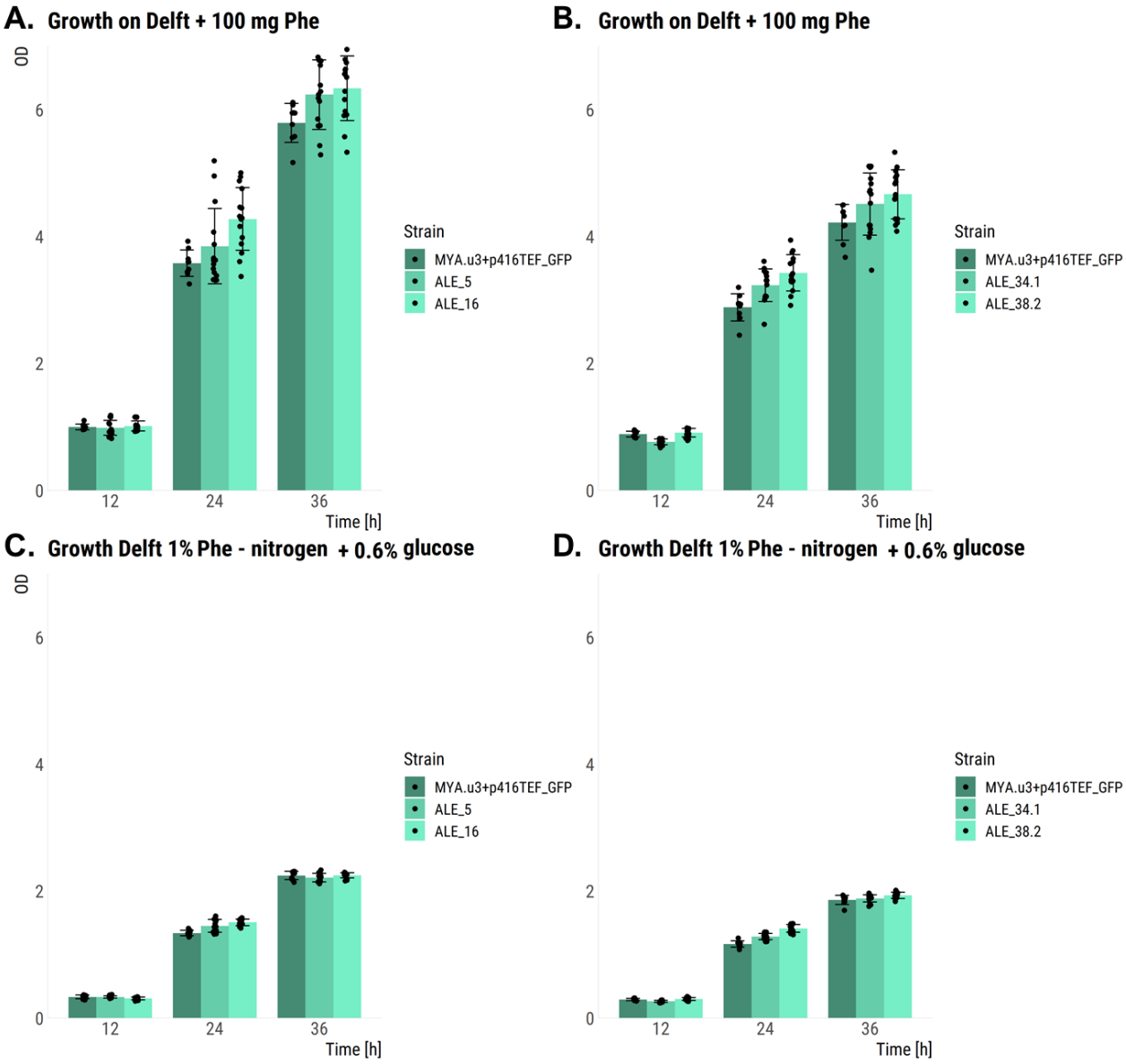

Figure 8. Growth of ALE strains in two different media, Delft +100 mg Phe (2% glucose 0.01% Phe) and Delft 1% Phe + 0.6% glucose – nitrogen. ODs were plotted together with standard deviations for each strain at three different time points, representing glucose phase, ethanol phase, and stationary phase. Strains were grown on microtiter plates in a growth profiler at 30 °C at 200 rpm. 5 strain replicates and 3 assay replicates were used. P-values for OD comparison between strains is presented in Supplementary Tables S5 and S6.

Although there was little difference in growth of strains looking at the continuous growth curves, there were consistent trends in growth on Delft 1% Phe + 0.6% glucose – nitrogen that differentiated from when strains were grown on Delft +100mg Phe (**Supplementary Fig. S2 & S3**). These differences were mainly a weak diauxic shift, the diauxic shift being the shift from glycolysis to anaerobic utilization of ethanol, as well as a slight decrease in OD after the stationary phase was reached (**Supplementary Fig. S2 & S3;** Galdieri et al., 2010).

The initial intention was to compare strains both in terms of growth and Phe consumption, but the data from the Phe assay could not be analyzed due to low quality of the data. As the areas of internal standard peaks were consistent throughout all samples, it is probable that the sample

preparation was the cause. Due to there being insufficient time to repeat the assay, these results were excluded. As no remarkable results were found in comparison of growth between strains, and no strain was identified with a significantly increased Phe consumption, evaluation of strains for new genomic targets was excluded from the project.

4.3 Evaluation of phenylalanine consumption and growth in engineered strains

4.3.1 Comparative growth

Continuous growth of engineered strains MYA.a9, MYA.aT, MYA.aT.a9, and MYA.a9PAL compared to control strain MYA, as well as growth of strains MYA.u3+p416TEF_PAL and MYA.u3+p416TEF_PAH with plasmid based expression compared to MYA.u3+p416TEF, was evaluated on two sets of media (**Supplementary Fig. S5 & S7**). The media used were Delft + 100 mg Phe (2% glucose 0.01% Phe) and Delft 0.6% glucose 1% Phe-nitrogen, representing a regular minimal medium supplemented with Phe and a medium with limited glucose and high Phe concentration. Although these strains were not engineered to grow better on Phe rich medium, evaluating growth on these media might give insight in eventual growth impairment or increased toxicity of Phe caused by the modifications. MYA.aT.a9 strain replicate 2 was removed from analysis as the strain appeared to be an outlier based on that the strains grew to less than half the OD of the other two strain replicates (**Supplementary Fig. S6**). Although MYA.aT.a9 strain replicate 3 did reach the same final OD as strain replicate 1, the growth trend of the strain diverged from that of all other engineered strains in terms of a more pronounced diauxic shift (**Supplementary Fig. S6**). As it could not be determined if strain replicate 1 or 3 represented the behavior that could be expected from any general strain with the same modifications, both strains were included in analysis. Although all three strains showed differences in growth trends, strain 1 and 3 were the most similar as they reached similar ODs in the stationary phase (**Supplementary Fig. S6**).

In order to better evaluate significance of results, data was extracted for time points 12, 24, and 36 h to represent the difference in ODs mid-glucose phase, mid-ethanol phase, and in the stationary phase (**Fig. 9**). For MYA-based strains, deletion of *ARO9* resulted in significantly increased OD compared to control across several timepoints and media (p-value<0.05; **Fig. 9 A, C; Supplementary Tab. S4).** All additional modifications tested, i.e. *PAL* integration in the *ARO9* genetic locus or P_{TEF1} knock-in in the P_{AGP1} genomic locus, did not result in improved growth compared to the control strain (**Fig. 9 A, C; Supplementary Tab. S4).** In MYA.u3+p416TEF based strains no consistent trends were found (**Fig. 9 B, D; Supplementary Tab. S5).**

Following the same trend as the previously evaluated ALE strains, all strains grown in Delft 1% Phe + 0.6% glucose – nitrogen had a weak diauxic shift and a slight decrease in OD after reaching the stationary phase (**Supplementary Fig. S4**).

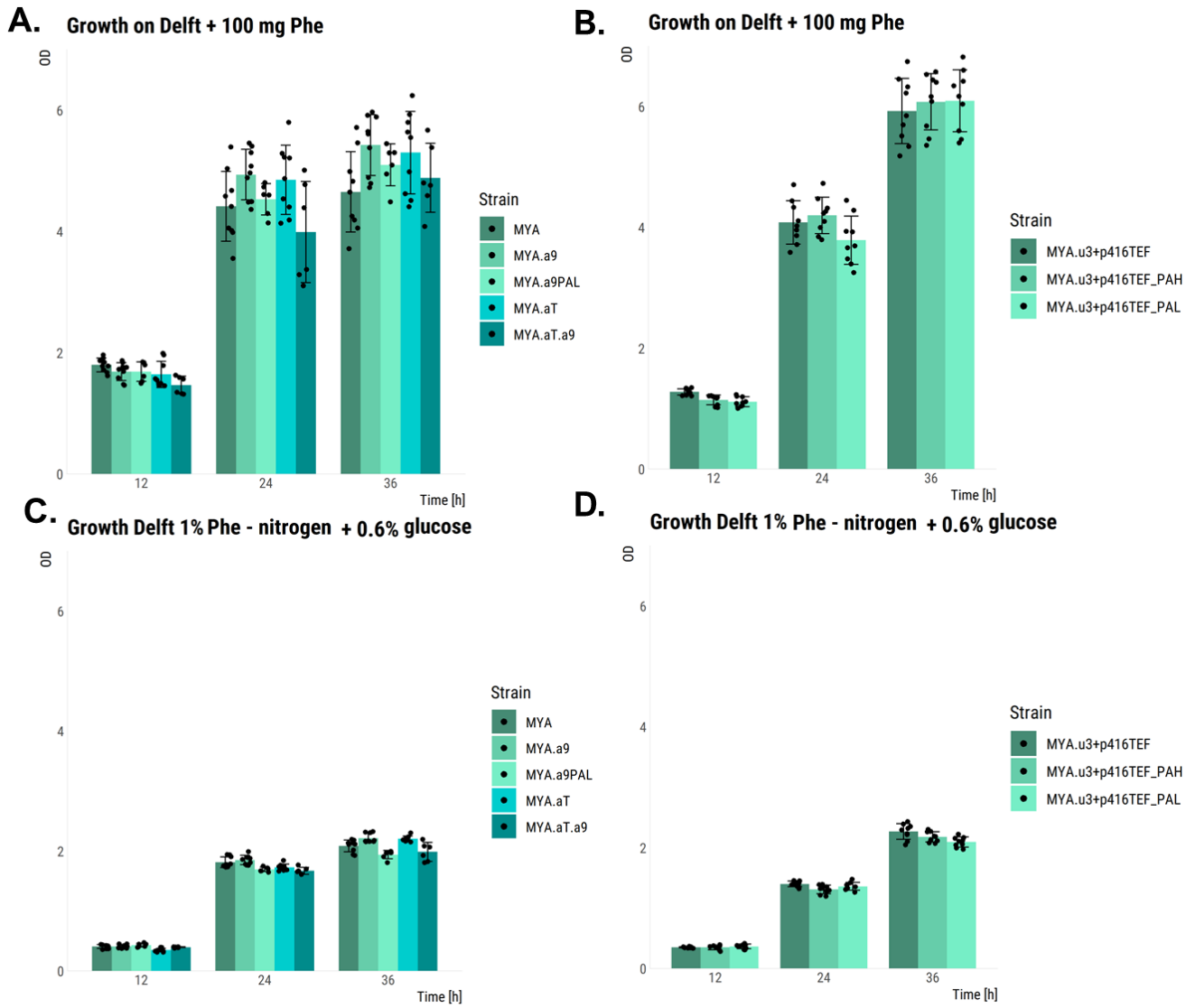

Figure 9. Growth of MYA-based strains and MYA.u3+p416TEF based strains on two different media, Delft +100 mg Phe (2% glucose 0.01% Phe) and Delft 1% Phe + 0.6% glucose – nitrogen. ODs were plotted together with standard deviations for each strain at three different time points, representing mid-glucose phase, mid-ethanol phase, and stationary phase. Strains were grown on microtiter plates in a growth profiler at 30 °C at 200 rpm. 3 strain replicates and 3 assay replicates were used. P-values for OD comparison between strains is presented in Supplementary Tables S7 and S8.

4.3.2 Comparison of phenylalanine consumption between strains

In order to see if the Phe consumption was increased in the engineered strains compared to the controls, strains were grown in Delft media supplemented with 100 mg Phe and sampled after 12 and 24 h. LC-MS/MS was applied for evaluation of the phenylalanine concentrations in the samples, and concentrations were normalized for OD at time of sampling to enable comparison per the number of cells and not just by the total amount consumed (**Fig. 10**).

Strains MYA.a9PAL, MYA.aT, and MYA.a9.aT consumed significantly more Phe compared to the control strain after 12 h of cultivation, based on the OD-normalized values (p-values<0.05; **Fig. 8 A**). Moreover, comparing strain MYA.a9PAL to MYA.a9 at the same timepoint we also found an increased Phe consumption based on the OD normalized values (**Fig. 8 A**). Looking at normalized consumption after 24 h instead, no strain was significantly

better compared to the control. However, at 24 h strain MYA.a9PAL and MYA.aT.a9 consumed significantly more Phe compared to MYA.a9, and MYA.aT.a9 consumed significantly more Phe compared to MYA.a9PAL (p-values<0.05; **Fig. 8 A**). Although stain MYA.a9PAL consumed the most Phe based on the OD normalized values, this strain consumed significantly less Phe based on the non-normalized consumption (p-value<0.05; **Fig. 8 C**). Comparing the OD normalized Phe consumption of all the MYA-based strains after 12 respectively 24 h of growth, it is evident that more Phe was consumed per OD the first 12 h compared to the last 12 h (**Figure 8 A**). While the consumption per OD at 24 h averaged at around 120 µM/OD for all strains, the least consumed amount at 12 h was around 160 µM/OD and the most about 210 µM/OD. The reason for this is unclear.

Examining the non-normalized Phe consumption respectively the OD at sampling, over half of the total Phe is consumed during the first 12 h of growth, while much of the increase in OD occurs during the last 12 h (**Fig. 8 C, E**). As an example, approximately 260 µM Phe was consumed and the OD reached 2.2 in the first 12 hours of growth of the MYA strain, while a total of 460 µM Phe had been consumed and the OD reached 4.5 after 24h (**Fig. 8 C, E**). One outlier strain replicate, strain MYA.aT.a9 number 2, was removed from analysis as this strain consistently, within assay replicates, diverged from the other strain replicates in terms of the amount of Phe consumed per OD (**Supplementary Fig. S7**).

In the plasmid-based strains, the Phe consumption, OD-normalized and non-normalized, was significantly increased for strain MYA.u3+p416TEF_PAL compared to the control after 12 h (p-value<0.05; **Fig. 8 B**). This was the only strain for which the increase in consumption was found to be significant in the non-normalized values as well, both at 12 and 24 h (p-value<0.05; **Fig. 8 D; Supplementary Tab. S6**). Furthermore, this was the only plasmid-based strain on pair with Phe consumption of the MYA-based strains, but only at 24 h. All plasmid-based strains grew slower compared to the MYA-based strains, possible due to the *URA3* gene form *S. cerevisiae* in the plasmid not being able to fully replace the function of the deleted *URA3* gene in *S. boulardii* (**Fig. 8 F**). The lower ODs caused normalized values for Phe consumption of plasmid-based strains to exceed those of the MYA-based strains, although the total Phe consumption was lower for all except MYA.u3+p416TEF_PAL (**Fig. 8 F**).

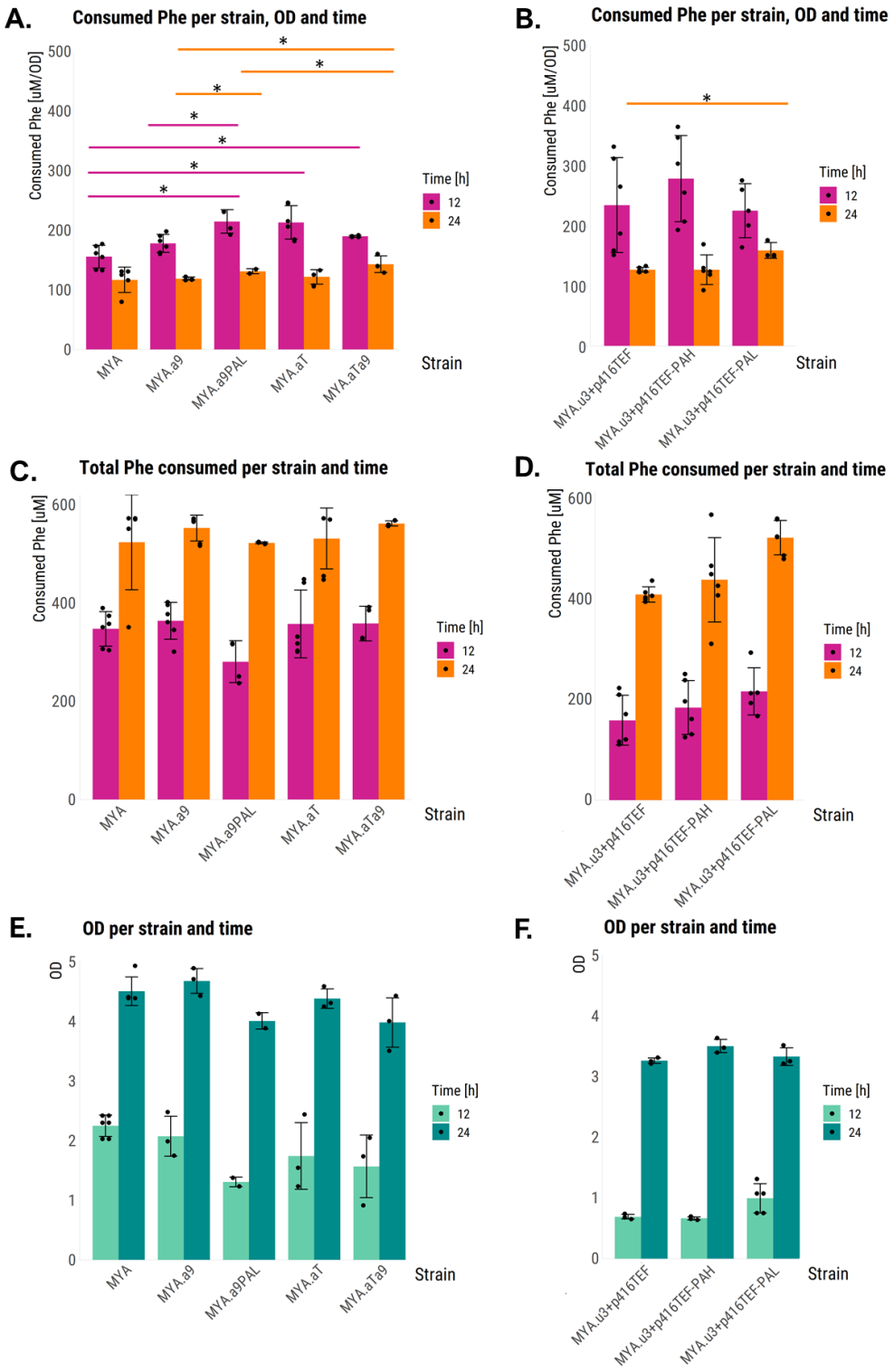

Figure 10. Plots containing results from evaluation of Phe consumption of each strain. Three transformation replicates were used for all strains, except for MYA.a9PAL for which only one transformation replicate was available. Two strain replicates were evaluated instead for this strain. Samples and ODs where collected at 12 and 24 h after inoculation of cultures, and phenylalanine levels were evaluated through LC-MS/MS with 2 assay replicates. Phe concentrations in the media was also evaluated, to allow for calculation of consumption. (**A**) Consumed Phe normalized per OD in MYA-based strains, (**B**) consumed Phe normalized per OD in MYA.u3+p416TEF based strains, (**C**) total consumed Phe in MYA-based strains, (**D**) total consumed Phe in MYA.u3+p416TEF based strains, (**E**) OD at time of measurement in MYA-based strains, (**F**) OD at time of measurement in MYA.u3+p416TEF based strains.

4.4 Metagenomic mycobiome analysis

4.4.1 Analysis of metagenomic data in the iHMP IBDMDB dataset

After initial processing of data as described in the methods section, we obtained a matrix containing the relative abundance of fungi in each sample. We proceeded to perform dimensionality reduction on the data using Multi-Dimensional Scaling (MDS), in order to visualize global differences in the mycobiota composition between healthy controls, UC, and CD patients (**Fig. 11**). We found that the majority of patients overlapped in terms of composition, although a subset of UC and CD patients clustered completely separately from the healthy controls. When performing hierarchical clustering of the dataset we again identified minor shifts in composition amongst patients (**Fig. 12**). As global compositional shifts between patient groups were minimal, we proceeded to extract individual differentially abundant fungi. Our analysis identified 18 differentially abundant fungi between each patient group ($p < 0.05$, unadjusted two-sided Kruskal-Wallis test; **Fig. 13**). Of these *Brettanomyces anomalus, Candida albicans, Candida tropicalis, Candida sp. LDI48194, Malassezia globosa, Pichia kudruavzevii, Penicillium roqueforti, Amauroascus niger, Clavaria fumosa,* and *Melampsora pinirorqua* were enriched in patients with IBD. *Penicillium biforme* and *Klyveromyces marxianus* enrichment were unique to UC patients and a lower abundance of *Alternata alternata* was unique to CD patients. Enrichment of *Agaricus bisporus.var bisporus, Chrysosporium queenslandicum,* and *Debaryomyces hansenii* was unique to healthy controls.

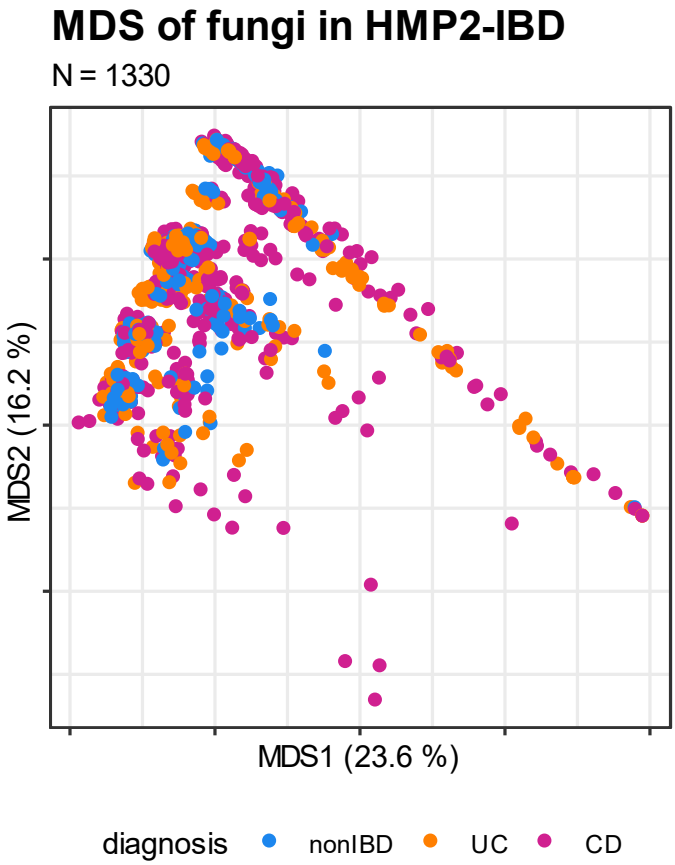

Figure 11. MDS plot of fungal abundances amongst patients in the iHMP-IBD cohort. Euclidean distance was used to calculate distances between pairs of patients.

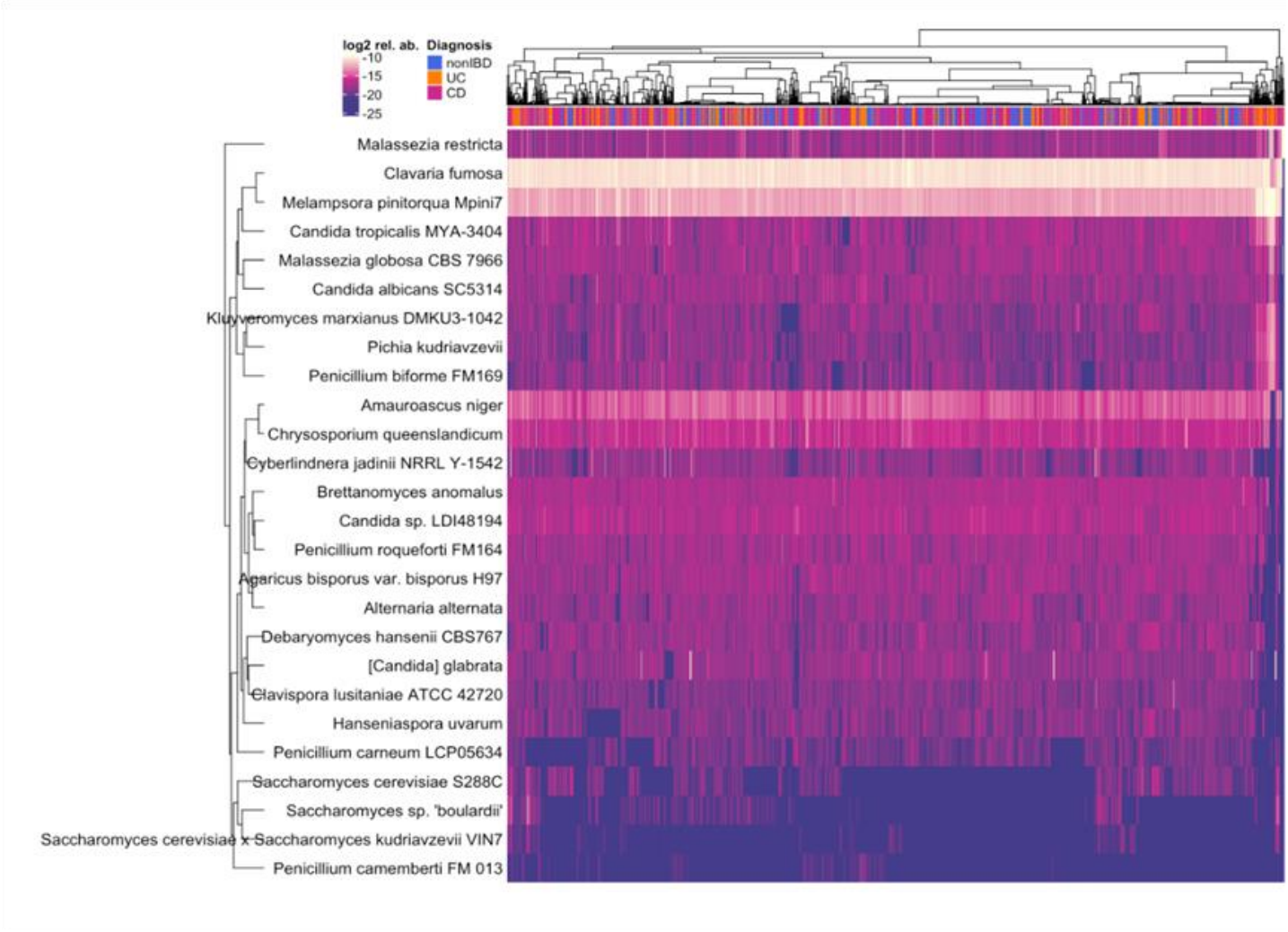

Figure 12. Heatmap of fungal abundance per patient and sample. Clustering was based on the Pearson correlation coefficient.

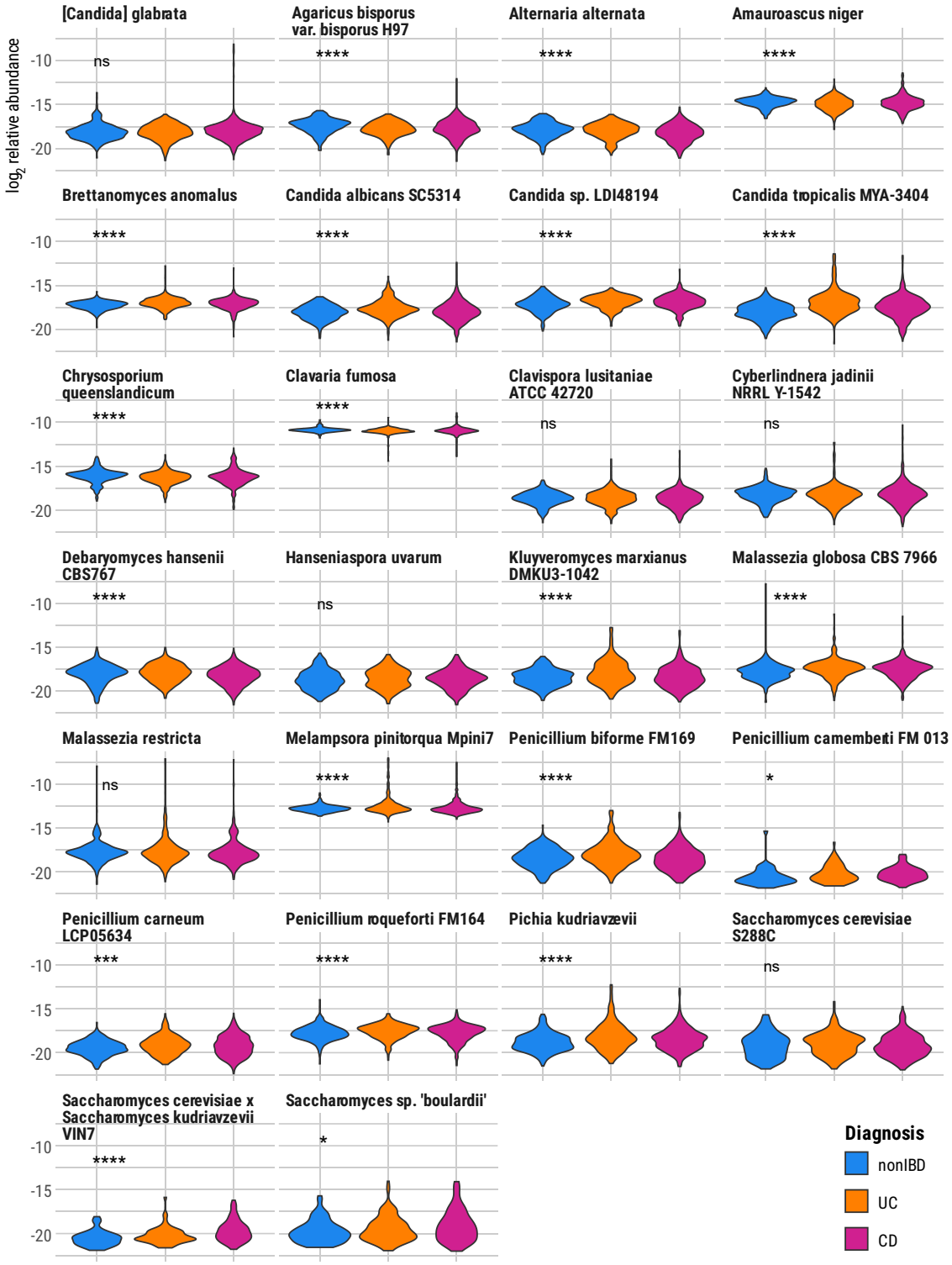

Figure 13. Violin plot of the total abundance of each fungus across samples for each condition. Significance was evaluated using the unadjusted two-sided Kruskal-Wallis test (* p<0.05, ** p<0.01, *** p<0.001, and **** p<0.0001).

5 Discussion

This project consisted of two parts, the main project being development of a *S. boulardii* strain able to consume high amounts of Phe for treatment of PKU, and the second project being identification of mycobiota compositions associated with IBD in metagenomic datasets. In the main project, the initial task was establishment of *S. boulardii* as a microorganism in the lab. Although engineering was successful for multiple targets using protocols developed for *S. cerevisiae,* the overall efficiency remained low even with optimization of the transformation protocol. Multiple targets were never engineered successfully, i.e. *ARO8, PHA2,* and *HIS3.* One reason to this may be that *S. boulardii* is very sensitive to Geneticin selection, which was found recently, which was the main selection applied in Cas9-mediated engineering during this project (Durmusoglu et al., 2020). This explains the increased efficiency of *ARO8* deletion when a plasmid with *KIURA* marker was used instead, as this would allow a larger number of cells to survive the selection. Furthermore, with uracil auxotrophy being a less harsh selection method in *S. boulardii* it is also possible that Cas9 and the gRNA were better expressed compared to when Geneticin was used, as the environment does not force cells to focus resources on survival. The toxicity of KanMX in *S. boulardii* also explains the increased efficiency of Cas9-mediated knockouts when overnight transformation was implemented, as the overnight recovery allowed for expression of both KanMX and Cas9 before selection pressure was applied. In continued engineering it is recommended to use an auxotrophic marker or alternatively NatMX, since the latter has been shown less toxic to the yeast compared to KanMX (Durmusoglu et al., 2020). Another factor that could contribute to the difference in efficiency using the different Cas9 plasmids, is that pCfB2312 is a low copy number plasmid while pKIURA_ECas9 is a high copy number palsmid. As the same promoters/terminators were used for expression of the gRNA and Cas9 in both pasmids, the higher plamid copy number of pKIURA_ECas9 should result in an overall higher expression of Cas9 and gRNA in the cells which could increase the efficiency. As indicated in the literature, the diploidy of *S. boulardii* sometimes caused one genetic copy of a gene targeted with Cas9 to remain, while the other genetic copy was successfully modified (**Fig. 5 C**; Durmusoglu et al., 2020)**.** This was very clear in the case of *AGP1* knock-in in Figure 5C**,** whereas in Figure 5D lane 6 and Figure 6 lane 2 clear bands indicating successfully implemented genomic modification were present as well as low intensity bands at the fragment length expected of the unedited gene. This indicates that a genomic copy may remain of the original gene, but the comparatively low intensity of the unedited bands compared to all other bands on the gel makes it difficult to draw a conclusion. The difference in efficiency of varying genomic targets has also been observed previously in *S. boulardii* (Durmusoglu et al., 2020). A possible explanation could be eventual small differences in the genomic sequences of *S. cerevisiae* and in the *S. bouladii bicodex* reference genomes, both used for gRNA design in this project, compared to the genomic sequence of MYA-796. This could result absence of some gRNA sites in MYA-796 that were present in *S. cerevisiae* and in *S. bouladii bicodex*. Such an event could be avoided by targeting multiple sites within a gene at the same time. This was not the case in engineering of *ARO8* however, as the gene was deleted successfully when a different selection marker was used. For deletion of genes causing auxotrophy, such as deletion of *HIS3* and *PAH2,* plating on media supplemented with higher concentrations of the respective amino acid might improve efficiency.

Although overnight recovery improved the efficiency of Cas9-mediated knockouts when the KanMX marker was used it was not applied when expression plasmids with the KIURA marker were transformed into the yeast. The reason for this was that the efficiency of transformation was sufficient already. Moreover, the potential drawbacks of overnight recovery are that strains may get the time to replicate and that plasmids may be dropped due to no selection pressure

being applied. The incubation temperature used in transformation of *S. boulardii* was altered from 37 °C to 30 °C partway through the project based on recommendation of Alex Hedin, who works with the same strain of *S. boulardii*. Whether or not this improved transformation efficiency is inconclusive.

In evaluation of strains in terms of Phe consumption, several of the engineered strains were found to have a significantly increased consumption of Phe compared to the control strain based on OD normalized values after 12 h of growth (**Fig. 10 A**). One of the strains with the highest consumption of Phe after 12 h was MYA.a9PAL, but looking at the total, non-normalized, consumption of Phe the strain consumed significantly less Phe compared to all other strains (p-value<0.05; **Fig. 10 C, Supplementary Tab. S5)**. The difference in the results are caused by the strain growing to a lower OD compared to the other strains (**Fig. 10 E**). However, the same strain was not found to have grown to a lower OD compared to the other strains in the evaluation of OD after 12, 24 and 36 h (**Fig. 9 A, C; Supplementary Tab. S5**). The difference in the results from these separate experiments in terms of OD of MYA.a9PAL after 12 h of growth on Delft+100 mg Phe may be caused by the difference in growth conditions. ODs associated with Phe consumption were measured in shake flask cultures, where oxygenation of media is more efficient compared to when strains are grown on microtiter plates as in the growth evaluation experiment. Alternatively, the difference could be caused by variation in inoculated concentration if MYA.a9PAL strains were inoculated at a lower OD unintentionally in the Phe assay. This is the reason for why the normalized values are the focus of Phe consumption evaluation. MYA.aT also had a large in consumption of Phe compared to the control.

ARO9 deletion did not cause evident toxicity and resulted in significantly increased OD on Delft+100mg after 26 and 36 h, and *AGP1* overexpression did not significantly impact OD, the combination of these modifications seemed to result in toxicity. These results are inconclusive however, due to that all biological strain replicates of MYA.aT.a9 behaved differently (**Supplementary Fig. S5**). Assuming that both of the analyzed MYA.aT.a9 strains, strain replicates 1 and 3, were representative for the implemented modifications, the observed toxicity may have been caused by that the absence of Aro9p combined with increased import of Phe and Tyr. As the reaction catalyzed by Aro9p is reversible it is able to both consume and produce Phe and Tyr, meaning that deletion of *ARO9* would remove one pathway through which Phe and Tyr could be processed when *AGP1* is overexpressed (Wu et al., 2020). Phe has also previously been found to cause toxicity when *AGP1* is overexpressed in *S. cerevisiae*, but based on the results of growth evaluation of MYA.aT in Phe rich media, overexpression of *AGP1* alone did not cause toxicity in *S. boulardii* (**Supplementary Tab. S4;** S.J. Ruiz et al. 2017).

The MYA.u3+p416TEF_PAL strain had a significantly higher consumption of Phe both with and without normalization for OD. No increase in Phe consumption was observed in the strain carrying p416TEF_PAH. This was expected as it was discovered halfway through the project that PAH requires the co-enzyme BH4 to function which is not produced by the yeast (Al Hafid and Christodoulou, 2015). As no heterologous genes were expressed for production of this co-enzyme, PAH was not expected to function. Since PAH could no longer releave the Tyr auxotrophy, an alternative strategy of *PHA2* deletion was constructed. While the initial idea was to achieve auxotrophy by deletion of both *ARO8* and *ARO9*, deletion of *PHA2* would result in Phe auxotrophy while allowing continued activity of Aro8p and Aro9p, and thereby two more pathways in which Phe could be consumed (Wu et al., 2020).

Although a slight increase in Phe consumption was observed in the engineered strains, with a total consumption of 440 µM at the most after 12 h, these cannot compete with the already

developed *E. coli* strains for treatment of PKU which is able to consume 1 mM Phe in 60 min (Isabella et al., 2018). In the development of this *E. coli* strain, the transporter for Phe uptake proved to be the most important (Isabella et al., 2018). Based on this, screening of different transporters for uptake of Phe could be a promising way to continue development oa a yeast competitor. Based on the results of the currently implemented modifications, expression of *PAL* seems to be important for increased Phe consumption. This is based on the OD normalized consumption (**Fig. 10**). Moreover, the strain MYA.u3+p416TEF_PAL which had plasmid-based expression of *PAL* was the only strain in which both normalized and non-normalized consumption was significantly increased at 24h (**Fig. 10)**. Based on this, multiple genomic integrations may be a promising approach to further increase Phe consumption. In evaluation of Phe consumption, it was also found that the OD normalized consumption of Phe was higher after 12 h of growth compared to after 24 h. This may stem from the fact that the *TEF1* promoter was used for all constitutive expression, and that this promoter is most active during the glucose phase of growth (Maury et al., 2018). This explanation still fails in strains which only have knockouts of genes, but evaluation of alternative promoters could still be applied to see if a better alternative could be found.

In the evaluation of growth of the ALE strains, multiple of the evolved strains were found to have a significantly increased OD after 24 h compared to the control (**Fig. 8; Supplementary Tab. S2 & S3**). The improvement was slight however, and none of the strains were able to grow on Phe as a carbon source, which was the overall aim of the ALE. These results are not surprising however, as *S. boulardii* does not have a preexisting pathway for growth on solely Phe. The duration of the ALE usually has to be long for development of new functions, but a successful evolution is still not guaranteed as it depends on what functions are missing. Even when a new function is added by implementation of a non-native pathway through genetic engineering, an extended time of ALE may be required to achieve full functionality. An example of this was published in Cell last year, where Shmuel Gleizer and colleagues engineered *E. coli* to express a non-native Calvin cycle for utilization of carbon dioxide as a carbon source (Gleizer et al., 2019). Although the pathway was functional, approximately 200 days of ALE was needed to achieve complete trophic mode change (Gleizer et al., 2019). Another possible reason to why *S. boulardii* was unable to utilize Phe as a carbon source is an insufficient or incorrectly applied evolutionary pressure.

In the metagenomic analysis of fungi present in healthy versus patients with IBD, small but significant differences were found between fungal strains and conditions. Although the differences in abundances are small, the results are shown to be significant through the large number of samples included in the analysis, resulting in a high power of data. Some of the identified fungi, including the opportunistic pathogen *Candida albicans, Candida tropicalis,* and *Kluyveromyces marxianus* have previously been implicated in IBD (Chehoud et al., 2015; Lewis et al., 2015; Sokol et al., 2017). This implies functionality of the developed pipeline for fungal identification. Multiple fungi that had not previously been identified in association with IBD were found to be significantly increased in abundance in IBD as well, such as *Malassezia globosa and Pichia kudriavzevii.* Of these *M. globosa* is associated with skin disease, and is closely related to *M. restricta* that has been implicated as a promoter of pancreatic cancer (Ashbee and Evans, 2002; Aykut et al., 2019). *P. kudriavzevii* has not previously been associated with IBD to our knowledge, and the species has been proposed as a probiotic in prevention of colorectal cancer together with *Kluyveromyces marxianus* (Rahbar Saadat et al., 2020). Several species were found to be more abundant in healthy controls as well, including *Debaromyces hansenii, Clavira fumosa,* and *Agaricus bisporus.* Of these *D. hansenii* has been implicated in production of anti-*Candida* mycotoxins and has previously been found to be

inversely correlated with the abundance of *Candida* present in the gut, but it has not previously been implicated to prevent IBD (Huseyin et al., 2017). *C. fumosa* and *A. bisporus* are both mushrooms, but while *A. bisporus* is also known as the champignon mushroom, commonly used in various cuisines across the world which could explain presence of its DNA, *C. fumosa* is not included in any cuisine to our knowledge. Moreover, *C. fumosa* is far more abundant compared to other fungal species detected, indicating that the reads mapping to the genome of this fungus might be misclassified bacterial reads. This is also the case for *Amauroascus niger* and *Melampsora pinirorqua.* In order to detect and exclude eventual misclassified species, manual curation will be performed where reads mapping to each detected fungus will be run through BLAST again, this time including databases for bacterial reads as well. Any misclassified fungal species will be excluded from the fungal database in further analysis, as misclassification indicates a low quality of the reference genome used to construct the database. To further analyze results, we will look deeper into eventual disease associations of all true positive fungi. Moreover, linear mixed-effect models will be constructed, where factors such as age, gender, diet, and what patent samples were collected from will be considered. This will give a better overview of results, as multiple factors that may influence the results other than the fungi present will be excluded.

In this thesis, the first steps were taken towards the aim of creating a *S. boulardii* strain able to consume large quantities of Phe. Although the currently developed strains cannot compete with the bacterial live biotherapeutics developed for treatment of PKU, there are several ways of improving efficiency of consumption. Therefore, whether or not *S. boulardii* is a suitable competitor as a live biotherapeutic for treatment of PKU cannot be concluded yet. The aim of performing a metagenomic analysis of the mycobiota composition in patients with IBD was successful, although further analysis will be needed before any conclusions can be drawn regarding the significance of the data.

6 References

Agamennone, V., Krul, C.A.M., Rijkers, G., Kort, R., 2018. A practical guide for probiotics applied to the case of antibiotic-associated diarrhea in The Netherlands. BMC Gastroenterology 18, 103. https://doi.org/10.1186/s12876-018-0831-x

Al Hafid, N., Christodoulou, J., 2015. Phenylketonuria: a review of current and future treatments. Translational Pediatrics 4, 304-317–317.

Altmann, M., 2017. The Benefits of Saccharomyces boulardii. The Yeast Role in Medical Applications. https://doi.org/10.5772/intechopen.70591

Ashbee, H.R., Evans, E.G.V., 2002. Immunology of Diseases Associated with Malassezia Species. Clin Microbiol Rev 15, 21–57. https://doi.org/10.1128/CMR.15.1.21-57.2002

Aykut, B., Pushalkar, S., Chen, R., Li, Q., Abengozar, R., Kim, J.I., Shadaloey, S.A., Wu, D., Preiss, P., Verma, N., Guo, Y., Saxena, A., Vardhan, M., Diskin, B., Wang, W., Leinwand, J., Kurz, E., Kochen Rossi, J.A., Hundeyin, M., Zambrinis, C., Li, X., Saxena, D., Miller, G., 2019. The fungal mycobiome promotes pancreatic oncogenesis via activation of MBL. Nature 574, 264–267. https://doi.org/10.1038/s41586-019-1608-2

Baeshen, N.A., Baeshen, M.N., Sheikh, A., Bora, R.S., Ahmed, M.M.M., Ramadan, H.A.I., Saini, K.S., Redwan, E.M., 2014. Cell factories for insulin production. Microbial Cell Factories 13, 141. https://doi.org/10.1186/s12934-014-0141-0

Brunello, L., 2019. Gut microbiota transfer experiments in germ-free animals. Nature Research. https://doi.org/10.1038/d42859-019-00009-z

Chehoud, C., Albenberg, L.G., Judge, C., Hoffmann, C., Grunberg, S., Bittinger, K., Baldassano, R.N., Lewis, J.D., Bushman, F.D., Wu, G.D., 2015. A Fungal Signature in the Gut Microbiota of Pediatric Patients with Inflammatory Bowel Disease. Inflamm Bowel Dis 21, 1948–1956. https://doi.org/10.1097/MIB.0000000000000454

Clark, H., 2019. Culturing anaerobes. Nature Research. https://doi.org/10.1038/d42859-019-00007-1

Deena, B., 2018. FDA approves BioMarin rare metabolic disorder drug, shares rise. Reuters.

Donovan, P.D., Gonzalez, G., Higgins, D.G., Butler, G., Ito, K., 2018. Identification of fungi in shotgun metagenomics datasets. PLOS ONE 13, e0192898. https://doi.org/10.1371/journal.pone.0192898

Dragosits, M., Mattanovich, D., 2013. Adaptive laboratory evolution – principles and applications for biotechnology. Microb Cell Fact 12, 64. https://doi.org/10.1186/1475-2859-12-64

Durmusoglu, D., Al'Abri, I., Collins, S.P., Beisel, C., Crook, N., 2020. Establishing Probiotic *Saccharomyces boulardii* as a Model Organism for Synthesis and Delivery of Biomolecules (preprint). Synthetic Biology. https://doi.org/10.1101/2020.01.22.915389

Edwards-Ingram, L., Gitsham, P., Burton, N., Warhurst, G., Clarke, I., Hoyle, D., Oliver, S.G., Stateva, L., 2007. Genotypic and Physiological Characterization of Saccharomyces boulardii, the Probiotic Strain of Saccharomyces cerevisiae. Applied and Environmental Microbiology 73, 2458–2467. https://doi.org/10.1128/AEM.02201-06

Gajewski, J., Pavlovic, R., Fischer, M., Boles, E., Grininger, M., 2017. Engineering fungal de novo fatty acid synthesis for short chain fatty acid production. Nature Communications 8, 1–8. https://doi.org/10.1038/ncomms14650

Galdieri, L., Mehrotra, S., Yu, S., Vancura, A., 2010. Transcriptional Regulation in Yeast during Diauxic Shift and Stationary Phase. OMICS: A Journal of Integrative Biology 14, 629–638. https://doi.org/10.1089/omi.2010.0069

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

Gleizer, S., Ben-Nissan, R., Bar-On, Y.M., Antonovsky, N., Noor, E., Zohar, Y., Jona, G., Krieger, E., Shamshoum, M., Bar-Even, A., Milo, R., 2019. Conversion of Escherichia coli to Generate All Biomass Carbon from CO2. Cell 179, 1255-1263.e12. https://doi.org/10.1016/j.cell.2019.11.009

Hanley, W.B., 2004. Adult phenylketonuria. The American Journal of Medicine 117, 590–595. https://doi.org/10.1016/j.amjmed.2004.03.042

Hudson, L.E., McDermott, C.D., Stewart, T.P., Hudson, W.H., Rios, D., Fasken, M.B., Corbett, A.H., Lamb, T.J., 2016. Characterization of the Probiotic Yeast Saccharomyces boulardii in the Healthy Mucosal Immune System. PLoS ONE 11, e0153351. https://doi.org/10.1371/journal.pone.0153351

Huseyin, C.E., O'Toole, P.W., Cotter, P.D., Scanlan, P.D., 2017. Forgotten fungi—the gut mycobiome in human health and disease. FEMS Microbiol Rev 41, 479–511. https://doi.org/10.1093/femsre/fuw047

Isabella, V.M., Ha, B.N., Castillo, M.J., Lubkowicz, D.J., Rowe, S.E., Millet, Y.A., Anderson, C.L., Li, N., Fisher, A.B., West, K.A., Reeder, P.J., Momin, M.M., Bergeron, C.G., Guilmain, S.E., Miller, P.F., Kurtz, C.B., Falb, D., 2018. Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. Nat Biotechnol 36, 857–864. https://doi.org/10.1038/nbt.4222

Jovel, J., Patterson, J., Wang, W., Hotte, N., O'Keefe, S., Mitchel, T., Perry, T., Kao, D., Mason, A.L., Madsen, K.L., Wong, G.K.-S., 2016. Characterization of the Gut Microbiome Using 16S or Shotgun Metagenomics. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.00459

Kelesidis, T., Pothoulakis, C., 2012. Efficacy and safety of the probiotic *Saccharomyces boulardii* for the prevention and therapy of gastrointestinal disorders. Therap Adv Gastroenterol 5, 111–125. https://doi.org/10.1177/1756283X11428502

Khatri, I., Tomar, R., Ganesan, K., Prasad, G.S., Subramanian, S., 2017. Complete genome sequence and comparative genomics of the probiotic yeast Saccharomyces boulardii. Sci Rep 7, 371. https://doi.org/10.1038/s41598-017-00414-2

Lewis, J.D., Chen, E.Z., Baldassano, R.N., Otley, A.R., Griffiths, A.M., Lee, D., Bittinger, K., Bailey, A., Friedman, E.S., Hoffmann, C., Albenberg, L., Sinha, R., Compher, C., Gilroy, E., Nessel, L., Grant, A., Chehoud, C., Li, H., Wu, G.D., Bushman, F.D., 2015. Inflammation, Antibiotics, and Diet as Environmental Stressors of the Gut Microbiome in Pediatric Crohn's Disease. Cell Host & Microbe 18, 489–500. https://doi.org/10.1016/j.chom.2015.09.008

Lloyd-Price, J., Arze, C., Ananthakrishnan, A.N., Schirmer, M., Avila-Pacheco, J., Poon, T.W., Andrews, E., Ajami, N.J., Bonham, K.S., Brislawn, C.J., Casero, D., Courtney, H., Gonzalez, A., Graeber, T.G., Hall, A.B., Lake, K., Landers, C.J., Mallick, H., Plichta, D.R., Prasad, M., Rahnavard, G., Sauk, J., Shungin, D., Vázquez-Baeza, Y., White, R.A., Braun, J., Denson, L.A., Jansson, J.K., Knight, R., Kugathasan, S., McGovern, D.P.B., Petrosino, J.F., Stappenbeck, T.S., Winter, H.S., Clish, C.B., Franzosa, E.A., Vlamakis, H., Xavier, R.J., Huttenhower, C., 2019. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. Nature 569, 655–662. https://doi.org/10.1038/s41586-019-1237-9

Marusich, W.C., Jensen, R.A., Zamir, L.O., 1981. Induction of L-phenylalanine ammonia-lyase during utilization of phenylalanine as a carbon or nitrogen source in Rhodotorula glutinis. Journal of Bacteriology 146, 1013–1019. https://doi.org/10.1128/JB.146.3.1013-1019.1981

Maury, J., Kannan, S., Jensen, N.B., Öberg, F.K., Kildegaard, K.R., Forster, J., Nielsen, J., Workman, C.T., Borodina, I., 2018. Glucose-Dependent Promoters for Dynamic Regulation of Metabolic Pathways. Front. Bioeng. Biotechnol. 6. https://doi.org/10.3389/fbioe.2018.00063

McCullough, M.J., Clemons, K.V., McCusker, J.H., Stevens, D.A., 1998. Species Identification and Virulence Attributes ofSaccharomyces boulardii (nom. inval.). Journal of Clinical Microbiology 36, 2613–2617.

Miyoshi, J., Sofia, M.A., Pierre, J.F., 2018. The evidence for fungus in Crohn's disease pathogenesis. Clin J Gastroenterol 11, 449–456. https://doi.org/10.1007/s12328-018-0886-9

Moré, M.I., Swidsinski, A., 2015. Saccharomyces boulardii CNCM I-745 supports regeneration of the intestinal microbiota after diarrheic dysbiosis – a review. CEG 237. https://doi.org/10.2147/CEG.S85574

Mukherjee, P.K., Sendid, B., Hoarau, G., Colombel, J.-F., Poulain, D., Ghannoum, M.A., 2015. Mycobiota in gastrointestinal diseases. Nat Rev Gastroenterol Hepatol 12, 77–87. https://doi.org/10.1038/nrgastro.2014.188

Nyhan, W.L., Barshop, B.A. and Al-Aqeel, A.I., 2012. Gaucher disease. *ATLAS OF INHERITED METABOLIC DISEASES. third ed. London (UK): Hodder Arnold*, pp.698-707.

Orphanet: Phenylketonuria [WWW Document], n.d. URL https://www.orpha.net/consor/cgi-bin/OC_Exp.php?Lng=EN&Expert=716 (accessed 3.5.20).

Ozdemir, T., Fedorec, A.J.H., Danino, T., Barnes, C.P., 2018. Synthetic Biology and Engineered Live Biotherapeutics: Toward Increasing System Complexity. Cell Systems 7, 5–16. https://doi.org/10.1016/j.cels.2018.06.008

Pace, N.R., 1997. A Molecular View of Microbial Diversity and the Biosphere. Science 276, 734–740. https://doi.org/10.1126/science.276.5313.734

PALYNZIQ® (pegvaliase-pqpz) Injection for PKU, n.d. . BioMarin. URL https://www.biomarin.com/products/palynziq/ (accessed 3.25.20).

Pariente, N., 2019. A field is born. Nature Research. https://doi.org/10.1038/d42859-019-00006-2

Petersson, A., Almeida, J.R.M., Modig, T., Karhumaa, K., Hahn-Hägerdal, B., Gorwa-Grauslund, M.F., Lidén, G., 2006. A 5-hydroxymethyl furfural reducing enzyme encoded by the Saccharomyces cerevisiae ADH6 gene conveys HMF tolerance. Yeast 23, 455–464. https://doi.org/10.1002/yea.1370

Phenylketonuria. Genetic and Rare Diseases Information Center (GARD) [WWW Document], n.d. . Genetic and Rare Diseases Information Center (GARD) – an NCATS Program. URL https://rarediseases.info.nih.gov/diseases/7383/phenylketonuria (accessed 3.5.20).

Pietz, J., Kreis, R., Rupp, A., Mayatepek, E., Rating, D., Boesch, C., Bremer, H.J., 1999. Large neutral amino acids block phenylalanine transport into brain tissue in patients with phenylketonuria. J Clin Invest 103, 1169–1178. https://doi.org/10.1172/JCI5017

Powerful PKU control without a low-Phe diet requirement* [WWW Document], n.d. . Palynziq. URL https://www.palynziq.com/pku-treatment (accessed 4.15.20).

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J., Xu, J., Li, Shaochuan, Li, D., Cao, J., Wang, B., Liang, H., Zheng, H., Xie, Y., Tap, J., Lepage, P., Bertalan, M., Batto, J.-M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H., Yu, C., Li, Shengting, Jian, M., Zhou, Y., Li, Y., Zhang, X., Li, Songgang, Qin, N., Yang, H., Wang, Jian, Brunak, S., Doré, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P.,

Ehrlich, S.D., Wang, Jun, 2010. A human gut microbial gene catalogue established by metagenomic sequencing. Nature 464, 59–65. https://doi.org/10.1038/nature08821
Qin, J., Li, Y., Cai, Z., Li, Shenghui, Zhu, J., Zhang, F., Liang, S., Zhang, W., Guan, Y., Shen, D., Peng, Y., Zhang, D., Jie, Z., Wu, W., Qin, Y., Xue, W., Li, J., Han, L., Lu, D., Wu, P., Dai, Y., Sun, X., Li, Z., Tang, A., Zhong, S., Li, X., Chen, W., Xu, R., Wang, M., Feng, Q., Gong, M., Yu, J., Zhang, Y., Zhang, M., Hansen, T., Sanchez, G., Raes, J., Falony, G., Okuda, S., Almeida, M., LeChatelier, E., Renault, P., Pons, N., Batto, J.-M., Zhang, Z., Chen, H., Yang, R., Zheng, W., Li, Songgang, Yang, H., Wang, Jian, Ehrlich, S.D., Nielsen, R., Pedersen, O., Kristiansen, K., Wang, Jun, 2012. A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490, 55–60. https://doi.org/10.1038/nature11450
Rahbar Saadat, Y., Yari Khosroushahi, A., Movassaghpour, A.A., Talebi, M., Pourghassem Gargari, B., 2020. Modulatory role of exopolysaccharides of Kluyveromyces marxianus and Pichia kudriavzevii as probiotic yeasts from dairy products in human colon cancer cells. Journal of Functional Foods 64, 103675. https://doi.org/10.1016/j.jff.2019.103675
Regenberg, B., Kielland-Brandt, M.C., 1999. Substrate speci®city and gene expression of the amino-acid permeases in Saccharomyces cerevisiae. Journal of Amino Acids 12.
Rice, G.M., Steiner, R.D., 2016. Inborn Errors of Metabolism (Metabolic Disorders). Pediatrics in Review 37, 3–17. https://doi.org/10.1542/pir.2014-0122
Richard, M.L., Sokol, H., 2019. The gut mycobiota: insights into analysis, environmental interactions and role in gastrointestinal diseases. Nat Rev Gastroenterol Hepatol 16, 331–345. https://doi.org/10.1038/s41575-019-0121-2
Rooks, M.G., Veiga, P., Wardwell-Scott, L.H., Tickle, T., Segata, N., Michaud, M., Gallini, C.A., Beal, C., van Hylckama-Vlieg, J.E., Ballal, S.A., Morgan, X.C., Glickman, J.N., Gevers, D., Huttenhower, C., Garrett, W.S., 2014. Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. The ISME Journal 8, 1403–1417. https://doi.org/10.1038/ismej.2014.3
Ruiz, S.J., van 't Klooster, J.S., Bianchi, F., Poolman, B., 2017. Growth inhibition by amino acids in Saccharomyces cerevisiae. *bioRxiv*. https://doi.org/10.1101/222224
Sabu, C., Mufeedha, P., Pramod, K., 2019. Yeast-inspired drug delivery: biotechnology meets bioengineering and synthetic biology. Expert Opinion on Drug Delivery 16, 27–41. https://doi.org/10.1080/17425247.2019.1551874
Sáenz, D.A., Chianelli, M.S., Stella, C.A., 2014. L-Phenylalanine Transport in Saccharomyces cerevisiae: Participation of GAP1, BAP2, and AGP1. Journal of Amino Acids. https://doi.org/10.1155/2014/283962
Safety and Tolerability of SYNB1618 in Healthy Adult Volunteers and Adult Subjects With Phenylketonuria [WWW Document], n.d. . ClinicalTrials.gov. URL https://clinicaltrials.gov/ct2/show/NCT03516487 (accessed 4.15.20).
Sam, Q.H., Chang, M.W., Chai, L.Y.A., 2017. The Fungal Mycobiome and Its Interaction with Gut Bacteria in the Host. Int J Mol Sci 18. https://doi.org/10.3390/ijms18020330
Sokol, H., Leducq, V., Aschard, H., Pham, H.-P., Jegou, S., Landman, C., Cohen, D., Liguori, G., Bourrier, A., Nion-Larmurier, I., Cosnes, J., Seksik, P., Langella, P., Skurnik, D., Richard, M.L., Beaugerie, L., 2017. Fungal microbiota dysbiosis in IBD. Gut 66, 1039–1048. https://doi.org/10.1136/gutjnl-2015-310746
Spiller, R., Major, G., 2016. IBS and IBD — separate entities or on a spectrum? Nat Rev Gastroenterol Hepatol 13, 613–621. https://doi.org/10.1038/nrgastro.2016.141
Stone, L., 2019. Faecal microbiota transplantation for Clostridioides difficile infection. Nature Research. https://doi.org/10.1038/d42859-019-00008-0

Stovicek, V., Borodina, I., Forster, J., 2015. CRISPR–Cas system enables fast and simple genome editing of industrial Saccharomyces cerevisiae strains. Metabolic Engineering Communications 2, 13–22. https://doi.org/10.1016/j.meteno.2015.03.001

Tang, L., 2019. Sequence-based identification of human-associated microbiota. Nature Research. https://doi.org/10.1038/d42859-019-00011-5

Thompson, J.R., Register, E., Curotto, J., Kurtz, M., Kelly, R., 1998. An improved protocol for the preparation of yeast cells for transformation by electroporation. Yeast 14, 565–571. https://doi.org/10.1002/(SICI)1097-0061(19980430)14:6<565::AID-YEA251>3.0.CO;2-B

Valles-Colomer, M., Falony, G., Darzi, Y., Tigchelaar, E.F., Wang, J., Tito, R.Y., Schiweck, C., Kurilshikov, A., Joossens, M., Wijmenga, C., Claes, S., Van Oudenhove, L., Zhernakova, A., Vieira-Silva, S., Raes, J., 2019. The neuroactive potential of the human gut microbiota in quality of life and depression. Nat Microbiol 4, 623–632. https://doi.org/10.1038/s41564-018-0337-x

van Spronsen, F.J., 2010. Phenylketonuria: a 21st century perspective. Nature Reviews Endocrinology 6, 509–514. https://doi.org/10.1038/nrendo.2010.125

van Wegberg, A.M.J., MacDonald, A., Ahring, K., Bélanger-Quintana, A., Blau, N., Bosch, A.M., Burlina, A., Campistol, J., Feillet, F., Giżewska, M., Huijbregts, S.C., Kearney, S., Leuzzi, V., Maillot, F., Muntau, A.C., van Rijn, M., Trefz, F., Walter, J.H., van Spronsen, F.J., 2017. The complete European guidelines on phenylketonuria: diagnosis and treatment. Orphanet J Rare Dis 12. https://doi.org/10.1186/s13023-017-0685-2

Vliet, D. van, Bruinenberg, V.M., Mazzola, P.N., Faassen, M.H.J.R. van, Blaauw, P. de, Kema, I.P., Heiner-Fokkema, M.R., Anholt, R.D. van, Zee, E.A. van der, Spronsen, F.J. van, 2015. Large Neutral Amino Acid Supplementation Exerts Its Effect through Three Synergistic Mechanisms: Proof of Principle in Phenylketonuria Mice. PLOS ONE 10, e0143833. https://doi.org/10.1371/journal.pone.0143833

Wang, J., Jia, H., 2016. Metagenome-wide association studies: fine-mining the microbiome. Nature Reviews Microbiology 14, 508–522. https://doi.org/10.1038/nrmicro.2016.83

Williams, R.A., Mamotte, C.D., Burnett, J.R., 2008. Phenylketonuria: An Inborn Error of Phenylalanine Metabolism. Clin Biochem Rev 29, 31–41.

Wu, Y., Cai, M., Song, X., Li, Y., Wang, H., Mao, J., Liu, Q., Xu, H., Qiao, M., 2020. Comparative transcriptome analysis of genomic region deletion strain with enhanced l-tyrosine production in Saccharomyces cerevisiae. Biotechnol Lett 42, 453–460. https://doi.org/10.1007/s10529-019-02784-1

Zmora, N., Suez, J., Elinav, E., 2019. You are what you eat: diet, health and the gut microbiota. Nature Reviews Gastroenterology & Hepatology 16, 35–56. https://doi.org/10.1038/s41575-018-0061-2

Supplementary

Table S1. All primers used in this study are listed in this table. Lowercase sequence represents the binding part of the primers.

Primer name	Sequence
fw_SbMARK_PDR18	tggaagacatatggattgtggtgg
rv_SbMARK_PDR18	cttgccaactagctcaacgcc
fw_ScMARK_PGA3	ctcgttgccacccaaacttatc
rv_ScMARK_PGA3	ccatggagattgtgggtatgattac
ITS1	tccgtaggtgaacctgcg
ITS4	tcctccgcttattgatatgc
fw_gRNA_URA3	ATACGAAGTTATATTAAGGGTTGTCGACCTGCAGCGAAGCTTCAagacataaaaaac
fw_gRNAPlas_ARO8	CTCCGCAGTGAAAGATAAATGATCTTAGGTGATACTTTATACGAgtttttagagctagaaatagcaag
rv_gRNAPlas	GATCATTTATCTTTCACTGCGGAG
fw_gRNAPlas_PHA2	CTCCGCAGTGAAAGATAAATGATCACTCTTTTAAAGGTGACCTGgtttttagagctagaaatagcaag
fw_gRNAPlas_HIS3	CTCCGCAGTGAAAGATAAATGATCGGATGAGGCACTTTCCAGAGgttttagagctagaaatagcaag
fw_colPCR_URA3	tggttctggcgaggtattgg
rv_colPCR_URA3	atgtggctgtggtttcaggg
fw_colPCR_HIS3	accacttgccacctatcacc
rv_colPCR_HIS3	tggcctcctctagtacactc
fw_PHA2colPCR	catctgcgacaagcgtacag
rv_PHA2colPCR	cgtccaaatggaatggtctagag
fw_colPCR_AGP1	acgaaggtgtgcattctc
rv_colPCR_AGP1	tacaagatagaacccataatagc
fw_colPCR_ARO8	ttgcggattttcatatcacg
rv_colPCR_ARO8	tcacatttttcactctgcac
fw_colPCR_ARO9	cagagagaagttgaagaggag
rv_colPCR_ARO9	ttgaaagcgagtacggtg
fw_TEF1_AGP1don	TCCTTCGTGATTAGTTTCTTTTTGTTATTTTCCTCGTAATACTCATTTGTTTTACATACAcaaaatgtttctactccttttttactc
rv_CYC1_AGP1don	TATTTCTGTGGAGCTATTTTTCAAGTCTTTCAGTTCGTATAGAGACTTCGACGACGACATtttgtaattaaaacttagattagattgc
fw_PAL-bindTEF	TAGCAATCTAATCTAAGTTTTAATTACAAAatggctccatctctagaca
rv_PAL-bindCYC	AATGTAAGCGTGACATAACTAATTACATGAttatgccaacatttttagtagaacg
fw_PAH-bindTEF	AGCAATCTAATCTAAGTTTTAATTACAAAatgtctactgccgttttgg
rv_PAH-bindCYC	GCGTGACATAACTAATTACATGAttactttatttttgcaatgcgg

fw_TEF-ARO9don	ATATCCCAGTGTAAACATAAAAATACACACATACCACAATTACACTCTCTCATCGACTCAcaaaatgtttctactccttttttactc
rv_CYC-ARO9don	TACAAAAAACTACAATTTATATATGATATGCAAATAACACATAGGTTTAATCTTCACTGTgcaaattaaagccttcgagc
fw_TEF-ARO8don	CACTAAGTTAAACCCTGCAGTTGATACAGACATTGAATAGGACAACCGATCGTTACTATCcaaaatgtttctactccttttttactc
rv_CYC-ARO8don	TACAAAAATACGGACGTCCTTTTTTCACCTTATATATATTCTTCCAATGTATTTACCTCTgcaaattaaagccttcgagc
fw_CYC_PALPAH	TTGAACACTTCACAGATGATAGGGATTCGGgcaaattaaagccttcgagc
rv_TEF_PALPAH	ATACTCTATGACAGTCGACACAACACCGTCCGCcaaaatgtttctactccttttttactc
fw_p416_PALPAH	GCGGACGGTGTTGTGTCGACcTGTCATAGAGTATcaaaatgtttctactccttttttttactc
rv_p416_PALPAH	CCGAATCCCTATCATCTGTGAAGTGTTCAAagctccagctttgttcc

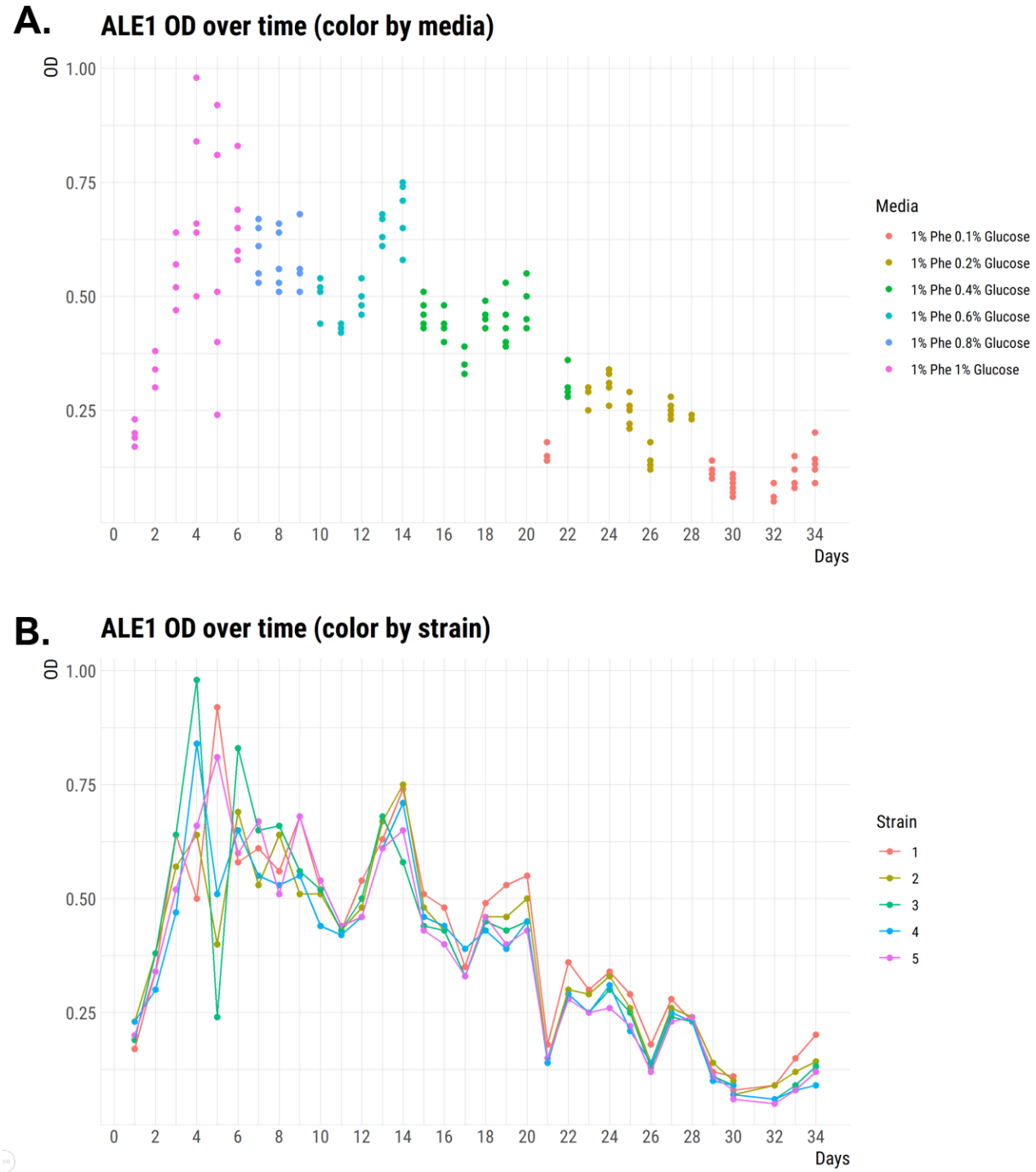

Figure S1. The first ALE was performed in 5 parallel shake flasks in media with a constant Phe level at 1% and a glucose level decreasing from 1% to 0.1% over time of the ALE. Strains were diluted to OD 0.05 in fresh media every 24 h, and glucose content of media was decreased when the OD at time of dilution had stabilized or started to increase. (A) shows the ODs measured after 24 h of growth colored by the media used. (B) shows the ODs colored by the ALE replicate.

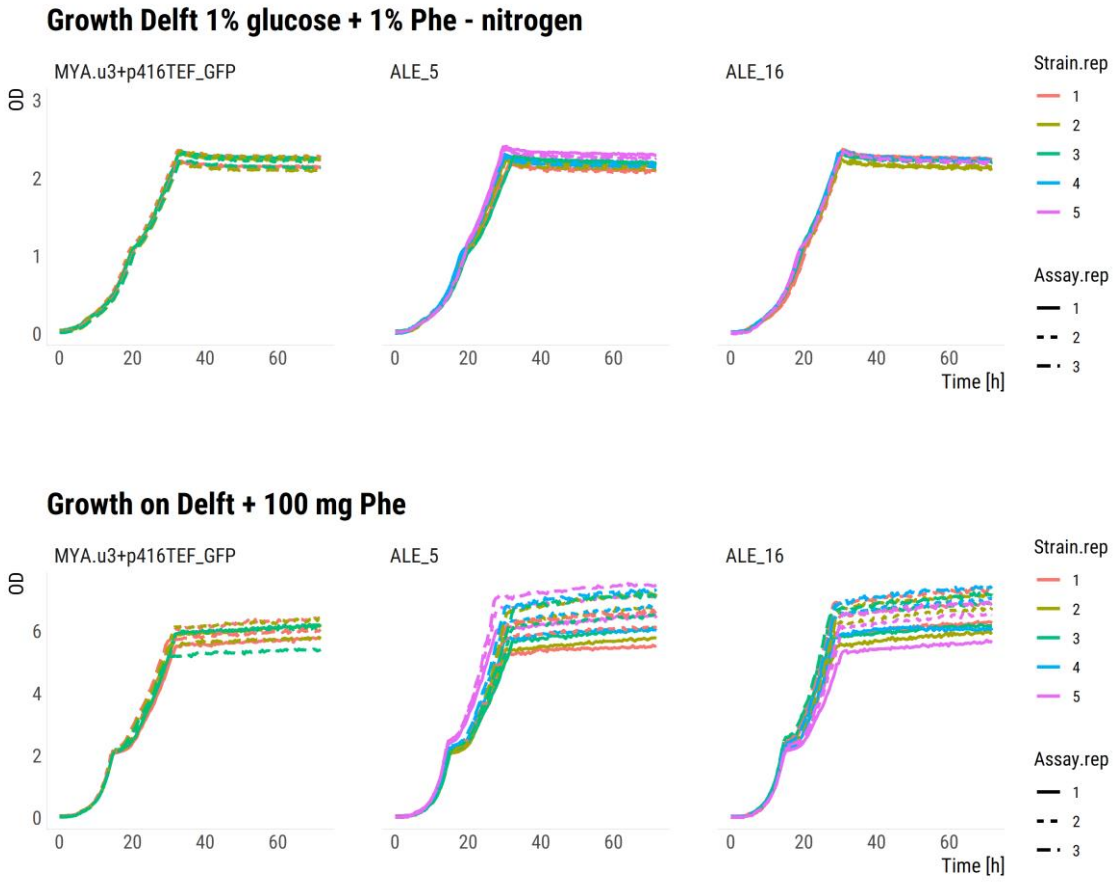

Figure S2. Continuous growth curves of ALE_5 and ALE_16 strains compared to control on two types of media. The line color indicates what transformation replicate the line belongs to and the line type shows what assay replicate the line represents.

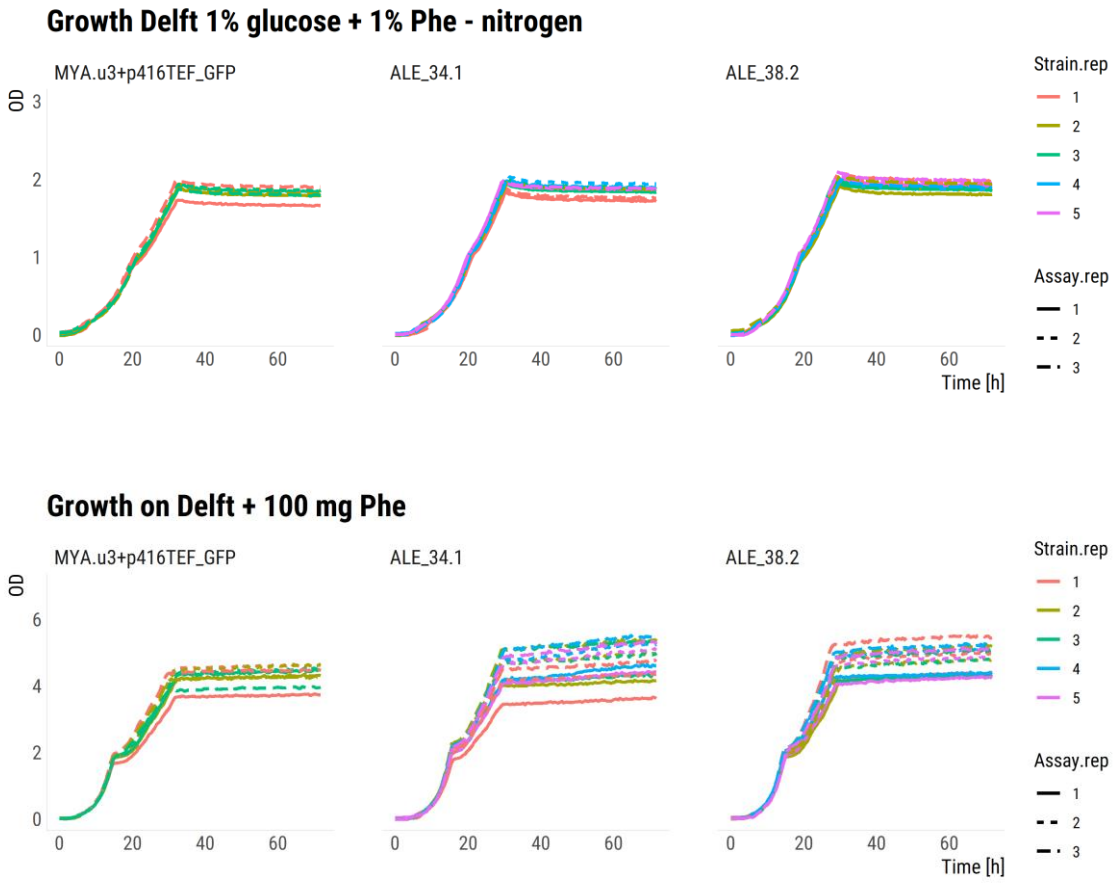

Figure S3. Continuous growth of ALE_34.1 and ALE_38.2 strains compared to control on two types of media. The line color indicates what transformation replicate the line belongs to and the line type shows what assay replicate the line represents.

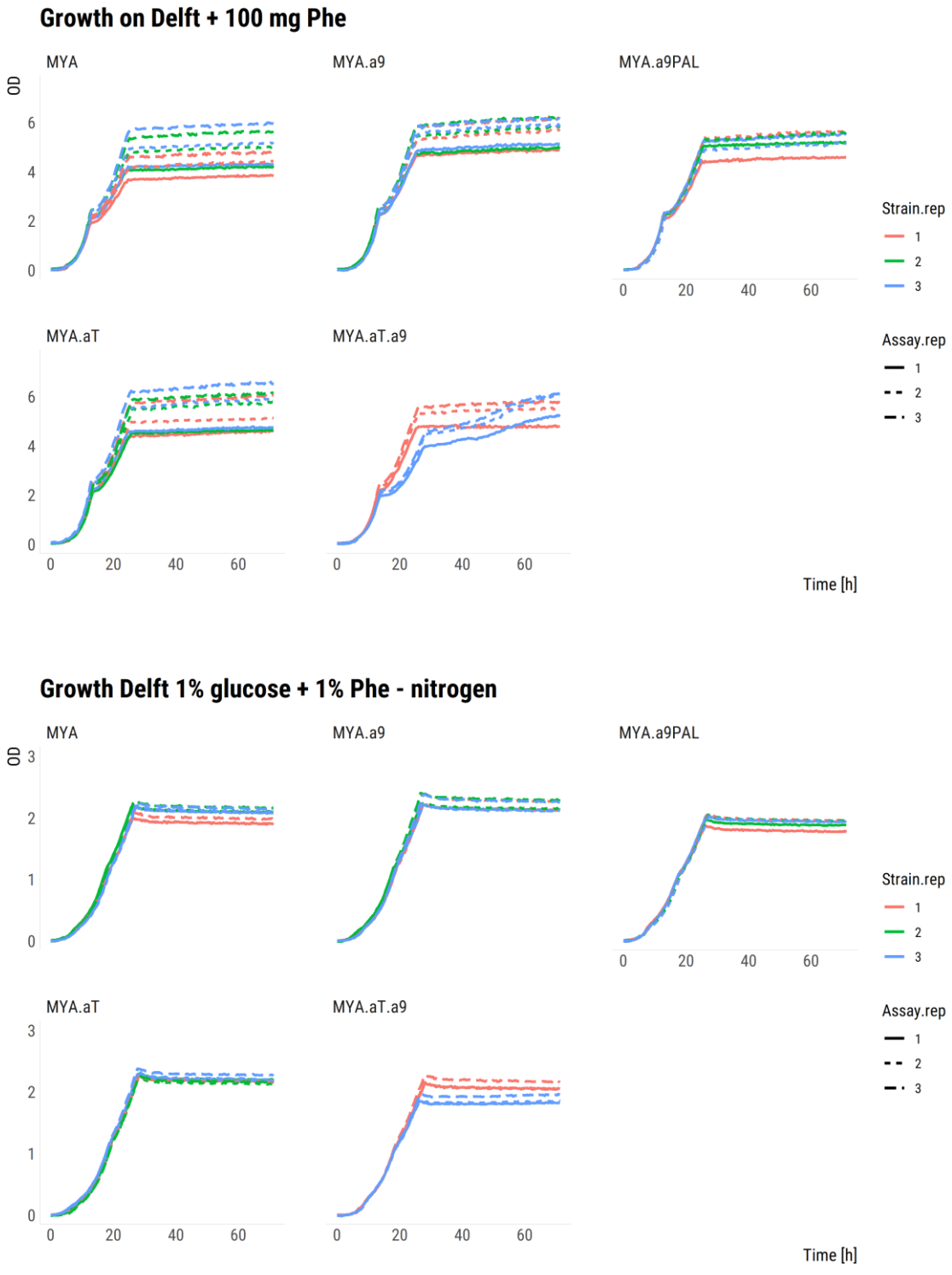

Figure S4. Continuous growth of all MYA based strains on two types of media. The line color shows what transformation replicate the line belongs to and the line type shows what assay replicate the line represents.

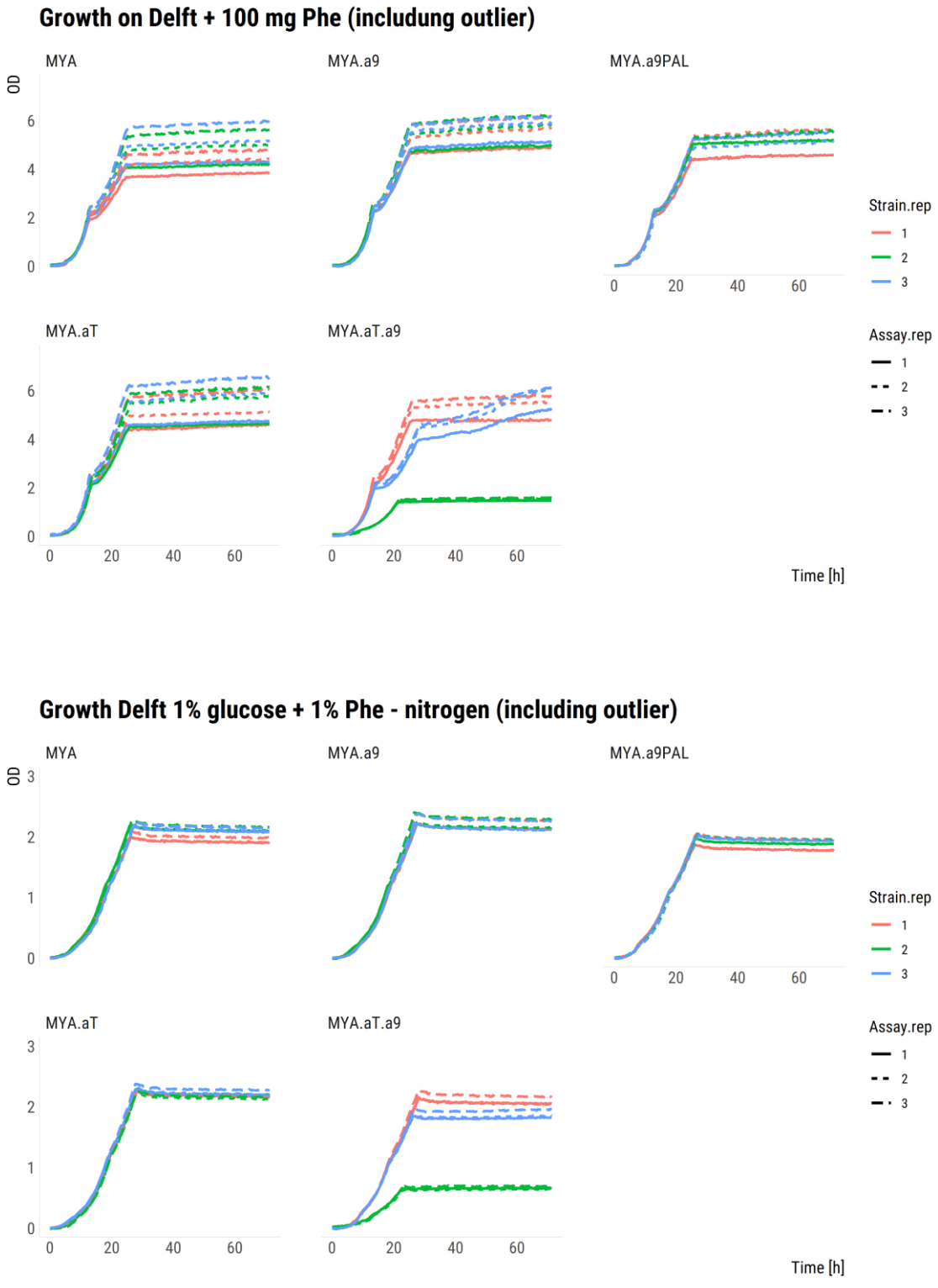

Figure S5. Continuous growth of all MYA based strains on two types of media, including MYA.a9.aT outlier. The line color shows what transformation replicate the line belongs to and the line type shows what assay replicate the line represents.

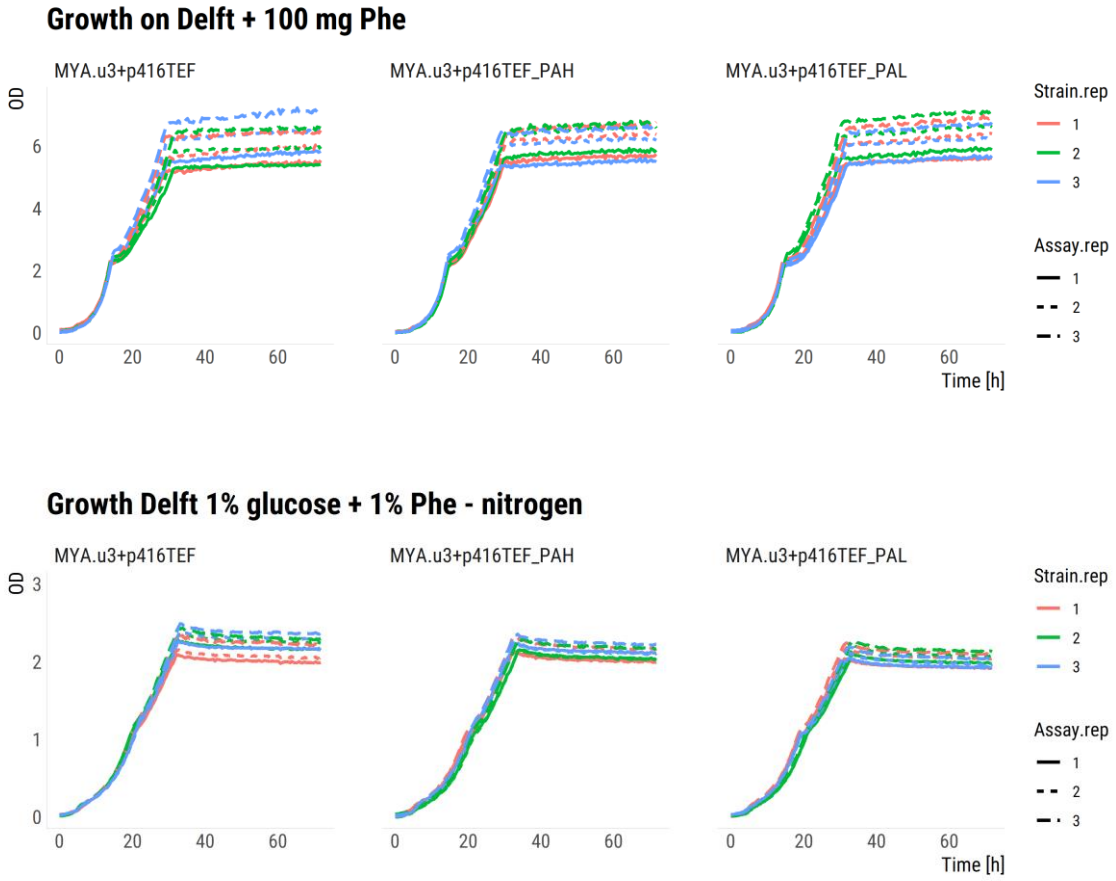

Figure S6. Continuous growth of MYA.u3 based strains on two types of media. The line color shows what transformation replicate the line belongs to and the line type shows what assay replicate the line represents.

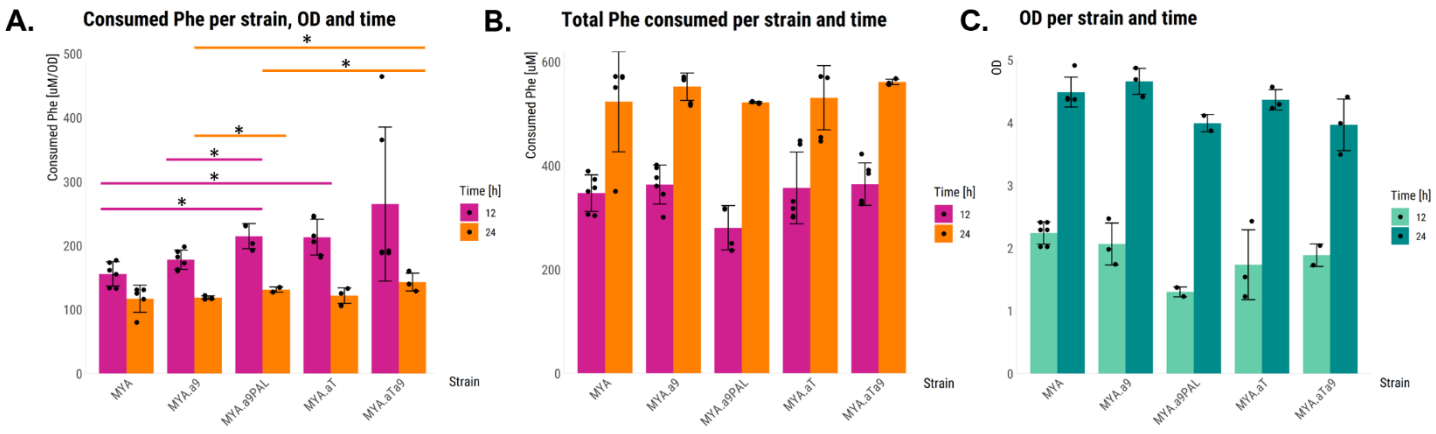

Figure S7. Plots containing results from evaluation of Phe consumption of each strain, including the MYA.aTa9 outlier. Three transformation replicates were used for all strains, except for MYA.a9PAL for which only one transformation replicate was available. Two strain replicates were evaluated instead for this strain. Samples and ODs where collected at 12 and 24 h after inoculation of cultures, and phenylalanine levels were evaluated through LC-MS/MS with 2 assay replicates. Phe concentrations in the media was also evaluated, to allow for calculation of consumption. (**A**) Consumed Phe normalized per OD in MYA-based strains, (**B**) total consumed Phe in MYA-based strains, (**C**) OD at time of measurement in MYA-based strains.

Table S2. P-values are listed for pairwise comparison of all strains in Figure 8 A and C, where ODs of ALE stains ALE_5 and ALE_16 are compared to OD of control strain MYA.u3+p416TEF_GFP on Delft +100 mg Phe media and Delft 1% Phe 0.6 % glucose media at three different timepoints. For each strain five transformation replicates and two assay replicates were used. Significant values, p <0.05, are marked in green.

Growth of strains ALE_5 & ALE_16 Delft+100mg Phe			Growth of strains ALE_5 & ALE_16 Delft_1% Phe 0.6%Glucose		
12 h			**12 h**		
Control strain	**Strain**	**p-value**	**Control strain**	**Strain**	**p-value**
MYA.u3+p416TEF_GFP	ALE_5	0.66042	MYA.u3+p416TEF_GFP	ALE_5	0.94894
MYA.u3+p416TEF_GFP	ALE_16	0.60596	MYA.u3+p416TEF_GFP	ALE_16	0.05987
ALE_5	ALE_16	0.44183	ALE_5	ALE_16	0.00585
24 h			**24 h**		
Control strain	**Strain**	**p-value**	**Control strain**	**Strain**	**p-value**
MYA.u3+p416TEF_GFP	ALE_5	0.12586	MYA.u3+p416TEF_GFP	ALE_5	0.00110
MYA.u3+p416TEF_GFP	ALE_16	0.00010	MYA.u3+p416TEF_GFP	ALE_16	4.57E-08
ALE_5	ALE_16	0.04061	ALE_5	ALE_16	0.07357
36 h			**36 h**		
Control strain	**Strain**	**p-value**	**Control strain**	**Strain**	**p-value**
MYA.u3+p416TEF_GFP	ALE_5	0.01791	MYA.u3+p416TEF_GFP	ALE_5	0.24825
MYA.u3+p416TEF_GFP	ALE_16	0.00342	MYA.u3+p416TEF_GFP	ALE_16	0.94685
ALE_5	ALE_16	0.61070	ALE_5	ALE_16	0.09513

Table S3. P-values are listed for pairwise comparison of all strains in Figure 8 B and D, where OD of ALE stains ALE_34.1 and ALE_38.2 are compared to OD of control strain MYA.u3+p416TEF_GFP on Delft +100 mg Phe media and Delft 1% Phe 0.6 % glucose media at three different timepoints. For each strain five transformation replicates and two assay replicates were used. Significant values, p <0.05, are marked in green.

Growth intermediate ALE strains Delft+100mg Phe			Growth interediate ALE strains Delft+100mg Phe		
12 h			**12 h**		
Control strain	**Strain**	**p-value**	**Control strain**	**Strain**	**p-value**
MYA.u3+p416TEF_GFP	ALE_34.1	0.000011529	MYA.u3+p416TEF_GFP	ALE_34.1	0.001006067
MYA.u3+p416TEF_GFP	ALE_38.2	0.324785300	MYA.u3+p416TEF_GFP	ALE_38.2	0.366881900
ALE_34.1	ALE_38.2	0.000000345	ALE_34.1	ALE_38.2	0.000027286
24 h			**24 h**		
Control strain	**Strain**	**p-value**	**Control strain**	**Strain**	**p-value**
MYA.u3+p416TEF_GFP	ALE_34.1	0.001935118	MYA.u3+p416TEF_GFP	ALE_34.1	0.000038382
MYA.u3+p416TEF_GFP	ALE_38.2	0.000031260	MYA.u3+p416TEF_GFP	ALE_38.2	0.000000002
ALE_34.1	ALE_38.2	0.059577080	ALE_34.1	ALE_38.2	0.000000810
36 h			**36 h**		
Control strain	**Strain**	**p-value**	**Control strain**	**Strain**	**p-value**
MYA.u3+p416TEF_GFP	ALE_34.1	0.079312919	MYA.u3+p416TEF_GFP	ALE_34.1	0.388874500
MYA.u3+p416TEF_GFP	ALE_38.2	0.004069923	MYA.u3+p416TEF_GFP	ALE_38.2	0.019859280
ALE_34.1	ALE_38.2	0.352426290	ALE_34.1	ALE_38.2	0.024890260

Table S4. P-values are listed for pairwise comparison of all strains in Figure 9 A and C, where OD of strains MYA.a9, MYA.a9PAL, MYA.aT, and MYA.aT.a9 was compared to OD of the control strain MYA on medias Delft +100 mg Phe and Delft 1% Phe 0.6 % glucose media. OD was compared at three different timepoints 12, 24 and 36 h. For each strain three transformation replicates and two assay replicates were used, with the exception of for MYA.a9PAL where two strain replicates were used instead. Significant values, p <0.05, are marked in green.

Growth MYA strains Delft+100mg Phe			Growth MYA strains Delft_1% Phe 0.6%Glucose		
	12 h			12 h	
Control strain	Strain	p-value	Control strain	Strain	p-value
MYA	MYA.a9	0.09454	MYA	MYA.a9	0.97028
MYA	MYA.a9PAL	0.18131	MYA	MYA.a9PAL	0.20555
MYA	MYA.aT	0.07889	MYA	MYA.aT	0.00098
MYA	MYA.aT.a9	0.00122	MYA	MYA.aT.a9	0.34503
MYA.a9	MYA.a9PAL	0.99947	MYA.a9	MYA.a9PAL	0.18556
MYA.a9	MYA.aT	0.60831	MYA.a9	MYA.aT	0.00078
MYA.a9	MYA.aT.a9	0.01561	MYA.a9	MYA.aT.a9	0.33901
MYA.a9PAL	MYA.aT	0.64678	MYA.a9PAL	MYA.aT	0.00044
MYA.a9PAL	MYA.aT.a9	0.03151	MYA.a9PAL	MYA.aT.a9	0.04634
MYA.aT	MYA.aT.a9	0.08187	MYA.aT	MYA.aT.a9	0.00122
	24 h			24 h	
Control strain	Strain	p-value	Control strain	Strain	p-value
MYA	MYA.a9	0.04325	MYA	MYA.a9	0.36214
MYA	MYA.a9PAL	0.60082	MYA	MYA.a9PAL	0.00389
MYA	MYA.aT	0.12439	MYA	MYA.aT	0.02634
MYA	MYA.aT.a9	0.30984	MYA	MYA.aT.a9	0.00231
MYA.a9	MYA.a9PAL	0.03643	MYA.a9	MYA.a9PAL	0.00019
MYA.a9	MYA.aT	0.72138	MYA.a9	MYA.aT	0.00138
MYA.a9	MYA.aT.a9	0.03805	MYA.a9	MYA.aT.a9	0.00019
MYA.a9PAL	MYA.aT	0.16684	MYA.a9PAL	MYA.aT	0.19144
MYA.a9PAL	MYA.aT.a9	0.18015	MYA.a9PAL	MYA.aT.a9	0.44652
MYA.aT	MYA.aT.a9	0.05759	MYA.aT	MYA.aT.a9	0.08896
	36 h			36 h	
Control strain	Strain	p-value	Control strain	Strain	p-value
MYA	MYA.a9	0.01383	MYA	MYA.a9	0.00664
MYA	MYA.a9PAL	0.11375	MYA	MYA.a9PAL	0.00545
MYA	MYA.aT	0.05745	MYA	MYA.aT	0.00687
MYA	MYA.aT.a9	0.48485	MYA	MYA.aT.a9	0.20618
MYA.a9	MYA.a9PAL	0.15850	MYA.a9	MYA.a9PAL	0.00001
MYA.a9	MYA.aT	0.66090	MYA.a9	MYA.aT	0.70402
MYA.a9	MYA.aT.a9	0.08782	MYA.a9	MYA.aT.a9	0.01356
MYA.a9PAL	MYA.aT	0.46372	MYA.a9PAL	MYA.aT	0.00005
MYA.a9PAL	MYA.aT.a9	0.45129	MYA.a9PAL	MYA.aT.a9	0.53543
MYA.aT	MYA.aT.a9	0.45129	MYA.aT	MYA.aT.a9	0.53543

Table S5. P-values are listed for pairwise comparison of all strains in Figure9 B and D, where OD of strains MYA.u3+p416TEF_PAL and MYA.u3+p416TEF_PAH was compared to the OD of control strain MYA.u3+p416TEF on medias Delft +100 mg Phe and Delft 1% Phe 0.6 % glucose media. OD was compared at three different timepoints 12, 24 and 36 h. For each strain three transformation replicates and two assay replicates were used. Significant values, p <0.05, are marked in green.

Growth p416TEF strains Delft+100mg Phe			Growth p416TEF strains Delft+100mg Phe		
12 h			12 h		
Control strain	Strain	p-value	Control strain	Strain	p-value
MYA.u3+p416TEF	MYA.u3+p416TEF_PAL	0.00021	MYA.u3+p416TEF	MYA.u3+p416TEF_PAL	0.27244
MYA.u3+p416TEF	MYA.u3+p416TEF_PAH	0.00124	MYA.u3+p416TEF	MYA.u3+p416TEF_PAH	0.87666
MYA.u3+p416TEF_PAL	MYA.u3+p416TEF_PAH	0.45764	MYA.u3+p416TEF_PAL	MYA.u3+p416TEF_PAH	0.34125
24 h			24 h		
Control strain	Strain	p-value	Control strain	Strain	p-value
MYA.u3+p416TEF	MYA.u3+p416TEF_PAL	0.11885	MYA.u3+p416TEF	MYA.u3+p416TEF_PAL	0.14336
MYA.u3+p416TEF	MYA.u3+p416TEF_PAH	0.47175	MYA.u3+p416TEF	MYA.u3+p416TEF_PAH	0.00543
MYA.u3+p416TEF_PAL	MYA.u3+p416TEF_PAH	0.02653	MYA.u3+p416TEF_PAL	MYA.u3+p416TEF_PAH	0.12492
36 h			36 h		
Control strain	Strain	p-value	Control strain	Strain	p-value
MYA.u3+p416TEF	MYA.u3+p416TEF_PAL	0.50719	MYA.u3+p416TEF	MYA.u3+p416TEF_PAL	0.00470
MYA.u3+p416TEF	MYA.u3+p416TEF_PAH	0.53132	MYA.u3+p416TEF	MYA.u3+p416TEF_PAH	0.11451
MYA.u3+p416TEF_PAL	MYA.u3+p416TEF_PAH	0.94605	MYA.u3+p416TEF_PAL	MYA.u3+p416TEF_PAH	0.04638

Table S6. P-values are listed for pairwise comparison of OD and total Phe consumption by all strains in Figure 10 C, D, E, and F. Strains were grown in triplicates on Delft +100 mg Phe media and sampled at 12 and 24 h. For each strain three transformation replicates and two assay replicates were used, with the exception of for MYA.a9PAL where two strain replicates were used instead. Significant values, $p < 0.05$, are marked in green.

12 h		
Control strain	**Strain**	**p-value**
MYA	MYA.a9	0.2864455
MYA	MYA.a9PAL	0.0000065
MYA	MYA.aT	0.0780287
MYA	MYA.aTa9	0.0194253
MYA.a9	MYA.a9PAL	0.0018306
MYA.a9	MYA.aTa9	0.3055809
MYA.aT	MYA.aTa9	0.5563354
MYA.u3+p416TEF	MYA.u3+p416TEF-PAL	0.04747559
MYA.u3+p416TEF	MYA.u3+p416TEF-PAH	0.19065368
24 h		
Control strain	**Strain**	**p-value**
MYA	MYA.a9	0.243665307
MYA	MYA.a9PAL	0.006601538
MYA	MYA.aT	0.359937714
MYA	MYA.aTa9	0.029756079
MYA.a9	MYA.a9PAL	0.000294556
MYA.a9	MYA.aTa9	0.007136427
MYA.aT	MYA.aTa9	0.06535346
MYA.u3+p416TEF	MYA.u3+p416TEF-PAL	0.325752016
MYA.u3+p416TEF	MYA.u3+p416TEF-PAH	0.001874079

.

DEPARTMENT OF BIOLOGY AND BIOLOGICAL ENGINEERING
DIVSION OF SYSTEMS AND SYNTHETIC BIOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY