Dissertation

submitted to the
Combined Faculties for the Natural Sciences and Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by
Dimitrios Konstantinidis, MSc
Born in Thessaloniki, Greece
Oral examination: July 3rd, 2020

Directed Evolution of Costly Metabolic Traits: use of microbial communities and metabolic modelling

Referees
Dr. Orsolya Barabas
Prof. Dr. Michael Knop

“Microorganisms will give you
anything you want,
if you know how to ask them.”
Kinichiro Sakaguchi

Acknowledgments

First and foremost, I would like to thank my supervisor Dr. Kiran Patil for accepting me as a PhD student and thus giving me the opportunity to become part of his group and to participate in exciting research projects. Kiran's imagination and ability to constantly come up with novel and meaningful experiments were catalytic for the development of my project and shaped my scientific beliefs and line of thought for the better. I am very thankful to Kiran for his guidance and his advice, but I am even more thankful about his patience, his goodness of heart and his optimism. The last along with some of his best jokes were crucial to help me overcome any setback that it might occur.

I would also like to thank the members of my thesis advisory committee Dr. Orsolya Barabas, Prof. Dr. Michael Knop, Dr. Nassos Typas and Dr. Ilka Bischoffs for investing precious time to give me advice and feedback and insightful input during the regular meetings of the committee. I am also thankful to Dr. Barabas and to Prof. Dr. Knop for accepting to examine my Thesis and to Prof. Dr. Rüdiger Hell for accepting to be part of the defence committee.

I want to take the opportunity to thank some of the people that I had the pleasure to meet and to get to know better throughout the duration of the PhD. First of all, I want to thank all the members of the Patil group, past and present for creating a pleasant working environment and each one of them teaching me something that will surely prove useful in the future. Through everyday lab life I interacted more with some members of the group and so I believe they deserve to be acknowledged in more detail.

I have to start by thanking Filipa Pereira, because when I first started my PhD she altruistically rushed to help me adapt both in and out of the lab, even though I am sure back then she didn't know what she was getting herself into. I am extremely thankful for her mentoring and for the fact that she stubbornly kept offering her help and advice, even when I wasn't the most cooperating person. I owe a lot to Filipa, because I believe that she helped me mature scientifically and she always had the best interest of my project at heart. Together we make a good team that can tackle any difficult. Apart from mentoring me in the lab, Filipa was always ready to offer her help in the real world, from offering me a warm home cooked meal to altering her dinner plans to help me transport furniture with her car. I hope I will be able to repay you one day for all these and much more.

The other person who patriotically tried to help me adjust when I first came to Germany was our group's fellow Greek PhD student (now Postdoc) Eleni Kafkia. She and her beautiful

dog Dione were always a comforting piece of Greece and eased my adjustment in Heidelberg. I will always keep in mind Eleni's most important piece of advice concerning life abroad "You need to lower your standards", which ranged from apartment hunting to purchase of pastries. Eleni is a great scientist who taught me a lot about metabolism and was always trying to help me with my experiments, even though our scientific interests differ. Last but not least, Eleni's loukoumades is the best way to end a dinner prepared by Filipa.

I also want to take the opportunity to thank Paula Jouhten and Natalia Gabrielli. I am thankful to Paula for entrusting one of her important project ideas to me and I hope I lived up to her expectations. It is always a pleasure to work with her and I would welcome the opportunity to collaborate again in the future. I would like to thank Natalia for improving my non-existent knowledge of the Spanish language and for her genuinely positive character. She always knows what to say to make everyone smile and I will always be amused by the thought that an ambulance parked outside of EMBL made her run to the lab, because I was changing a bottle of gas alone.

Of course, I cannot forget to thank Eva, Christoniki, Justyna, and Kristina for all the great times both at EMBL and during all the fun group activities. More specifically, I want to thank Eva for taking an interest in my project and spending many hours searching the available literature and recommending articles to me, while trying to make virtual sheep and cows grow together in harmony. I am very thankful to Christoniki for all the morning coffees, her genuine concern about everyone's wellbeing and for all the valuable tips when I first started tiptoeing into computational matters. I want to thank Justyna for keeping me in check in the lab and for making sure that I take proper precautions while handling all sorts of lab equipment, always with the appropriate amount of sarcasm in her comments. Finally, I want to thank Kristina for always being ready to assist me with one or another crazy idea in the lab and to try all sorts of experiments no matter how slim the chance of success might have seemed. She reminded me about the importance of getting excited by science and about acquiring new knowledge, despite the mishaps and the failures. I hope that she will keep this spark and the same thirst for knowledge in all her future steps.

I want to thank all the rest of close friends, past and present, that I made along the way. A special thank you to Akrivi Christidou, Dimitra Sakoula and Ioannis Kampouris who taught me so much during our common lab time in Greece and for all the fun experiences we shared since. Looking forward to see you both soon. I couldn't fail to mention my close friend and

housemate Philipp Hoess. We met at the beginning of our PhD life and through many football trips from Dortmund to Freiburg and all the fun Heidelberg moments we ended up sharing an apartment. I am grateful for all his help and comments during the preparation of this thesis and I can say with confidence that I couldn’t have been quarantined with better person during the spring of 2020.

Lastly, I want to thank my family for their love and support all these years. They have always done more than they could to see me happy and to encourage to follow my dreams. A big thanks to my mom who always made sure that I dress warm enough and that I eat properly, even though I was so far from home.

Summary

Evolution and adaptation through natural selection are cornerstone concepts of Biological sciences. The recent advances in the fields of Microbiology and Molecular Biology allowed scientists to introduce evolution in controlled laboratory settings. Adaptive laboratory evolution (ALE) has been successfully applied to better understand the effect of natural selection on individuals, as well as to obtain cells with improved phenotypic characteristics. In the majority of the reported cases, the characteristics that are targeted for improvement are related to biotechnological processes, aiming to create improved microbial strains for industrial applications. However, ALE is limited to growth-associated traits, such as substrate utilization and increased tolerance of compounds that inhibit growth.

The aim of this PhD thesis was to develop novel methodologies that could overcome the major bottleneck of ALE to enable the improvement of non-growth associated traits for non-genetically modified organism (GMO) biotechnological applications. In the first approach, small synthetic obligatory mutualistic communities were established. The design of a metabolic cross-feeding relationship between the species in the community couples the production of a target metabolite to the survival and proliferation of the community. Increased concentration of the target metabolite in the environment results in improved community fitness, despite of any potential production cost. Communities consisting of natural vitamin secreting lactic acid bacteria and engineered *Saccharomyces cerevisiae* were successfully evolved for the improved production of two different B group vitamins (riboflavin and folate). The isolated evolved overproducing bacterial strains can be used for the production of food with increased nutritional value.

The second approach described in this PhD thesis is a novel algorithm that uses genome-scale metabolic model simulations to identify the environmental conditions that will create selection pressure for the pathways involved in the production of a target compound. The computed chemical composition will be used as the environment (evolution niche) for ALE with straightforward growth selection. The resulting adapted metabolic network is expected to manifest the enhanced compound production when cells are switched back to their natural environment. As a proof-of-concept, we successfully applied this approach for the increased production of aroma compounds originating from the branched-chain or the aromatic amino acid pathways in wine yeast strains.

Together, the results of this thesis demonstrate that the developed methods can increase the precision of laboratory evolution and allow the selective production of fitness-costly metabolites. The phenotypic characteristics of both prokaryotes and eukaryotes could be improved, and the obtained strains hold potential for biotechnological applications, especially when the use of genetically engineered strains is restricted. Apart from the potential biotechnological applications, the designed laboratory evolution strategies can also be exploited to shed light on open questions about the physiology, the ecology and the social life of microbial species and communities.

Zusammenfassung

Evolution und Anpassung durch natürliche Selektion sind wichtige Grundprinzipien in den Biowissenschaften. Die jüngsten Fortschritte in den Bereichen Mikrobiologie und Molekularbiologie ermöglichten es den Wissenschaftlern, die Evolution in kontrollierten Laborumgebungen zu etablieren. Die adaptive Evolution im Labor (engl. *Adaptive Laboratory Evolution*, ALE) wurde erfolgreich angewandt, um die Auswirkungen der natürlichen Selektion auf individuelle Organismen besser zu verstehen und Zellen mit verbesserten phänotypischen Eigenschaften zu erhalten. Bei dieser Methode werden meist Merkmale verbessert, die für biotechnologische Prozesse wichtig sind. Dadurch entstehen mikrobielle Stämme, die in der Industrie Anwendung finden. Die ALE beschränkt sich jedoch auf Merkmale, die ein verbessertes Wachstum zur Folge haben, wie z. B. die Nutzung der verfügbaren Nährstoffe und eine erhöhte Toleranz gegenüber wachstumshemmenden Stoffen.

Das Ziel dieser Doktorarbeit war es, neue Methoden zu entwickeln, die die Beschränkung der ALE auf wachstumsassoziierte Merkmale überwinden. Dadurch wird es ermöglicht, den Phänotyp der in der Biotechnologie verwendeten mikrobiologischen Stämme zu verbessern, ohne dass sie dadurch zu gentechnisch veränderten Organismen (GVO) werden. Im ersten Kapitel wurden kleine synthetische Gemeinschaften aus Mikroorganismen etabliert, die auf mutualistische Symbiose angewiesen sind. Diese Gemeinschaften sind so entworfen, dass sich die Arten jeweils gegenseitig Stoffwechselprodukte zur Verfügung stellen, die für ihr Überleben und ihre Vermehrung notwendig sind. Eines dieser Stoffwechselprodukte ist der Zielmetabolit, dessen Produktion somit an das Wachstum der Gemeinschaft gekoppelt ist. Obwohl die Produktion des Zielmetaboliten möglicherweise mit einem hohen Energieverbrauch verbunden ist, führt seine Herstellung zu einer verbesserten Fitness der Gemeinschaft. In dieser Arbeit wurden Gemeinschaften, die aus natürlicherweise vitaminsezernierenden Milchsäurebakterien und manipulierter Bierhefe (*Saccharomyces cerevisiae*) bestehen, erfolgreich für die verbesserte Produktion von zwei verschiedenen Vitaminen der B-Gruppe (Riboflavin und Folat) evolviert. Die isolierten Milchsäurebakterien produzieren mehr Vitamine als am Anfang der Evolution und können daher für die Herstellung von Lebensmitteln mit erhöhtem Nährwert verwendet werden.

Die zweite Methode, die in dieser Doktorarbeit beschrieben wird, ist ein neuartiger Algorithmus, der Stoffwechselmodelle im Genom-Maßstab simuliert, um Umweltbedingungen zu identifizieren, die einen Selektionsdruck für die Produktion erwünschter Metaboliten darstellen. Die vom Algorithmus berechnete chemische Zusammensetzung wird als Rahmenbedingung (Nische für die Evolution) für ALE verwendet, wobei auf Basis des Wachstums selektiert wird. Es ist zu erwarten, dass der so angepasste Stoffwechsel die erwünschten Metaboliten auch produziert, wenn die Zellen wieder in ihre natürliche Umgebung zurückgebracht werden. Als Proof-of-Concept haben wir diesen Algorithmus angewandt, um in Weinhefen die Produktion von Aromastoffen zu erhöhen, die aus den Biosynthesewegen der verzweigtkettigen oder aromatischen Aminosäuren stammen.

Zusammengenommen deuten die Ergebnisse dieser Arbeit darauf hin, dass die entwickelten Methoden die Präzision der Evolution im Labor erhöhen und die selektive

Produktion von energieaufwändigen Metaboliten ermöglichen können. Darüber hinaus haben wir die phänotypischen Eigenschaften sowohl von Prokaryoten als auch von Eukaryoten verbessert, die in gentechnikfreier Biotechnologie oder solcher mit GVOs verwendet werden können. Abgesehen von den potenziellen biotechnologischen Anwendungen können die entworfenen Evolutionsstrategien auch genutzt werden, um offene Fragen zur Physiologie, Ökologie und zum sozialen Leben mikrobieller Arten und Gemeinschaften zu erforschen.

Table of Contents

Chapter 1: Introduction

Evolution and adaptation through natural selection are two of the most important concepts that define Biology. Traditionally scientists observed the effects of natural selection either by studying animal populations in nature or through the fossil archive of extinct species. The progress that occurred in scientific fields such as Microbiology and Molecular Biology allowed scientists to study evolution in laboratory conditions. Adaptive laboratory evolution (ALE) has been successfully applied to better understand the effect of natural selection on individuals, as well as to obtain cells with improved phenotypic characteristics. ALE is mostly used to improve microbial strains for biotechnological processes. However, use of ALE is limited by specific factors that are connected to cellular fitness. The following introduction explains the passage from traditional evolutionary studies to modern day use of evolution in laboratory and industrial settings, while at the end the objectives of the present thesis are presented.

1.1Evolution: from theory to practice

Mankind has always been fascinated by the variety of shapes and forms that life has adopted on Earth, which can be divided into the 5 kingdoms of life (Animalia, Plantae, Fungi, Protista, Archaea/Archaebacteria and Monera). Seemingly all environments that can support life from the deepest oceans to the highest mountains are the habitat of diverse species that interact with each other. This immense genetic variety of life is defined by the term biodiversity and the best effort to understand its origins and the mechanism behind the expansion of life was accomplished in 1859, when Charles Darwin published his work titled ‘On the origin of species’. Most importantly Darwin’s work provides a mean to explain how different organisms, in all their complexity and diversity share a common ancestry through natural processes (National Academy of Sciences. 2007). According to Darwin, these natural processes act through the natural selection of the most fit, or best adapted individuals, shaping the morphology, physiology, behaviour, and ecology of all the observed species. The effect of the theory of evolution on natural and life sciences can be summarised by the essay of the evolutionary biologist Theodosius Dobzhansky titled ‘Nothing in Biology Makes Sense Except in the Light of Evolution’ (Dobzhansky, 1973).

However, studying evolution is not always straightforward. Initially, the evolutionary studies were focused on higher eukaryotes, thus enhancing the belief that evolution is a slow, almost lethargic, process and the effects of natural selection are visible only over prolonged periods of time (Lenski, 2017). It was thought that the process of evolution may be observed best either through comparing fossil records and living species (van den Bergh *et al.*, 2018)

or in the absence of specific experimental approaches with theoretical studies (Conrad *et al.*, 2011). As a result, the field of evolution is brimming with abstract concepts that were thought to be impossible to test experimentally (Conrad *et al.*, 2011). Such comparative studies are still performed nowadays, enhanced with modern techniques like nucleic acid and protein sequencing, aiming to answer questions about the long-time evolutionary scale (van den Bergh *et al.*, 2018), but it was evident that a better study system was needed and it was meant to be found in the smallest life forms.

1.2 Microorganisms as model system for evolutionary biology

Darwin, as well as many other scientists, both evolutionary biologists and microbiologists, hypothesised early on that natural selection must act on microbes the same way as it was described for higher eukaryotes (Gupta, 2016). For example, the Dutch school of microbiologists in the 19th century focused on elucidating the evolutionary relationship among bacteria (Woese, 1987). Nevertheless, proving this hypothesis true and solving the phylogenetic relationships of microbes was challenging, mainly due to the difficulties that scientists were facing in accurately describing separate bacterial species and consequently, microbes were excluded from evolutionary studies (Koskella, 2015). Early classification systems for prokaryotic organisms focused on evaluating their morphology, physiology and biochemical characteristics. The disadvantage of such a system is that the chosen characteristics are versatile and may occur independently in distant organisms (Gupta, 2016). The revolution to the systematic classification of microorganisms was achieved through the field of molecular evolution. By capitalizing the technological advances of molecular biology, it was then possible to investigate potential evolutionary relationships through the comparison of conserved orthologous DNA regions present in various organisms. The DNA sequence of the small subunit ribosomal RNA (16S rRNA) is the best-known case of a gene being used for taxonomic classification of bacteria, and the use of which results in the separation of prokaryotes into two distinct taxa, bacteria and archaea (Gupta, 2016). Thus, the advances in microbiology and the clear systematic classification of those isolated species established the use of microbes in evolutionary studies and eased the transfer of the evolutionary process from the natural environment to the lab.

The major aim of experimental evolution is to introduce the evolutionary process in the controlled environment of a lab, to understand the mechanisms that drive forward natural selection. Of particular interest remains the mapping of the connection between genotype and phenotype, like the creation of the gene architecture of 500 traits in the yeast

Saccharomyces cerevisiae (Chen *et al.*, 2018). Alternatively, evolutionary studies are combined with biochemical and cell biological studies to understand key biological processes, such as the evolutionary history of organelles, like mitochondria and chloroplasts, which started off as endosymbiotic microorganisms (Richardson *et al.*, 2015). Another area of focus for experimental evolution is the elucidation of timescales and stages of transition, both at the level of individual cells as well as whole populations or diverse communities (Rainey *et al.*, 2017). Independently of the smaller research question of each study, the common goal of all the different studies of experimental evolution is to enhance the predictability of evolution. The outcome and the potential success of a study can be largely affected by the system being studied as well as the experimental design. In the case of evolutionary studies, the most common set up involves propagation of populations under defined conditions for a large number of generations and subsequent evaluation of the changes that accumulated over time (Kussel, 2013). Evolutionary studies that use microorganisms take advantage of the relatively short time periods that are needed to reach a large number of generations. Most microorganisms have a short generation time, ranging from 20 minutes in laboratory conditions for *Escherichia coli* (Gibson *et al.*, 2018) to 90 minutes for the yeast *Saccharomyces cerevisiae* (Salari and Salari, 2017) and thus the duration of the evolution experiment is in theory limited only by the amount of time the scientific team is willing to maintain the populations for (McDonald, 2019). Furthermore, this kind of studies offers the possibility for efficient reproducibility, as it is easy to handle multiple populations that all had the same initial starting point and were under the same environmental selection pressure. This increased reproducibility allows scientists to successfully identify and statistically analyse patterns of evolution (Bennet and Hughes, 2009). An important characteristic of this kind of studies is the fact that microbial populations can be stored for long periods of time in a frozen state and can be cultured again whenever it is necessary. This allows scientists to keep snapshots of the evolutionary process, which they can study separately and even compare with one another to evaluate the general changes in fitness, or even restart the experiment from a specific time point without starting the whole process from the beginning (McDonald, 2019).

Elucidation of the mechanism that drove adaptation and evolution forward has been enabled thanks to the rapid advances in the technology of nucleic acid sequencing. *De novo* DNA sequencing has become much cheaper and almost routine for many studies around the world. However, the price in combination with the amount of generated information still hinder the sequencing of multiple large genomes, like those of plants and animals and so

microbes with their small genome sizes offer one more advantage for evolutionary studies. Finally, the numerous powerful genetic tools that can be used to modify the genome and eventually the phenotype of microorganisms, as well as the natural richness of different species and strains that exhibit distinct biochemical capabilities (Kussel, 2013), allow scientists to design diverse studies to tackle interdisciplinary questions.

1.3 Adaptive Laboratory Evolution

Adaptive Laboratory Evolution (ALE) became an established laboratory method and has been used under different set ups to answer various scientific questions. One of the first recorded cases of ALE is the experiment performed by Reverend Dr. William Dallinger in 1887, who cultured protozoans for 7 years and was able to evolve them to tolerate temperatures up to 70 ^{o}C, but after the evolution process had passed they were unable to survive temperatures as low as 16 ^{o}C. This study was regarded an important demonstration of evolution at that time, suggesting that during the evolutionary process, a trade-off occurred between the ability to survive at low temperatures and the adaptation to higher temperatures (Bennett and Hughes, 2009). The term trade-off in laboratory evolution is used to describe the functional relationship between two different traits that occur during the evolution process when one trait that leads to an increase in fitness is countered by another trait that can potentially decrease the general fitness. Such an example is the increase of the number of offspring which has been associated with their decreased survival rate (Roff *et al.*, 2006). Whether the overall fitness will increase or not depends on the reproductive strategy of each species. Alternatively, trade-offs may emerge when an organism adapts under specific conditions, increasing its reproductive potential under these conditions, but simultaneously decreases under non-selective conditions, or the original environment from where the organism was isolated (Bennett and Hughes, 2009).

About one hundred years after the revolutionary experiment of Dallinger, Richard Lenski and his collaborators started one of the most famous long-term evolution experiments (Lenski *et al.*, 1991), which exhibited elegantly the process of adaptation through natural selection. In this experimental setup, twelve populations of *E. coli* are propagated in a glucose-limited minimal medium and a small percentage of the culture is diluted to a new one every day in fresh medium. This conceptually easy and straightforward experiment is still going on 32 years after its beginning and in 2017 around 66,000 generations had passed (Lenski, 2017). Lenski's study has helped elucidate various questions about the process of adaptation and the trajectories of fitness. One of the most important findings of this long-term evolution

experiment is that in the limited presence of glucose, the *E. coli* populations evolved to consume citrate, which is used as a buffer in various media recipes. The first mutations that led to a cit+ phenotype arose after 20,000 generations, a fact that highlights the importance of such long-term studies (Lenski, 2017).

Apart from the insight into the ecological and evolutionary questions that this study provided, it also helped with the development of novel technologies and tools. For a long time, the most common way to access evolution was to observe changes in fitness (Lang and Desai, 2014). The use of short-term parallel evolution, meaning the simultaneous evolution of a large number of identical initial populations for shorter time periods, can provide the statistical power that is required for the detection of subtler evolutionary changes (McDonald, 2019). Evolutionary adaptation depends on numerous factors, such as mutation rate, effect of single mutations, and the interaction between novel mutations. Hence, it is difficult to pinpoint only the beneficial mutations that arise in an evolving lineage (Tenaillon *et al.*, 2016). The effort to identify the mutations that arose during the evolution process, as well as to determine the cornerstone mutations that led to specific phenotypes, became a more approachable alternative thanks to novel sequencing technologies, as well as by using parallel evolution. However, under long-term evolution conditions, nonsynonymous mutations, intergenic mutations, insertions, and deletions are detected and interfere with one another, limiting even further the ability to choose only the beneficial mutations and further studying their impact on the phenotype. The use of parallel evolution can prove useful, as it offers the possibility to determine important mutations and eliminate others that occurred by stochastic processes, by selecting those that concurrently occur across different populations (Deathrage *et al.*, 2017). Efficiently mining and making sense of the large amount of data that is produced by sequencing can be compared to searching for a needle in the haystack and as a result, the development of new methods and tools became a necessity. One such tool is the pipeline Breseq, which was developed for mutation prediction in bacterial evolution studies (Deatherage and Barrick 2014). This kind of tools are not limited only to the evolutionary studies of single bacteria, but they can further be applied to characterise single isolates from genetic screens, as well as to identify undesired mutations that can occur through the processes of strain construction and the use of genome editing tools.

The most common experimental setup of ALE studies is the batch culture adaptation, during which an original strain or population is transferred in a serial manner under controlled environmental conditions, either in small liquid or solid (Petri dishes) cultures not requiring specialised equipment or sophisticated machinery (La Croix, 2017, Figure 1.1A). Through

the years, more sophisticated methodologies for ALE experiments have been described in the literature, such as chemostat cultures (Gresham and Hong, 2015, Figure 1.1B), microfluidic chips (Chen *et al.*, 2017, Figure 1.1C), emulsion cultures (Sjostrom *et al.*, 2014), selection for intensity of fluorescence, based on fluorescence-activated cell sorting (FACS) (Wildenberg *et al*, 2014), growth on alternating substrate (Sandberg *et al.*, 2017, Figure 1.1D) and propagation inside a living organism (Giraud *et al.*, 2001, Figure 1.1E). The selection of the appropriate methodology depends on the aims of the scientific questions asked and on the organism of choice.

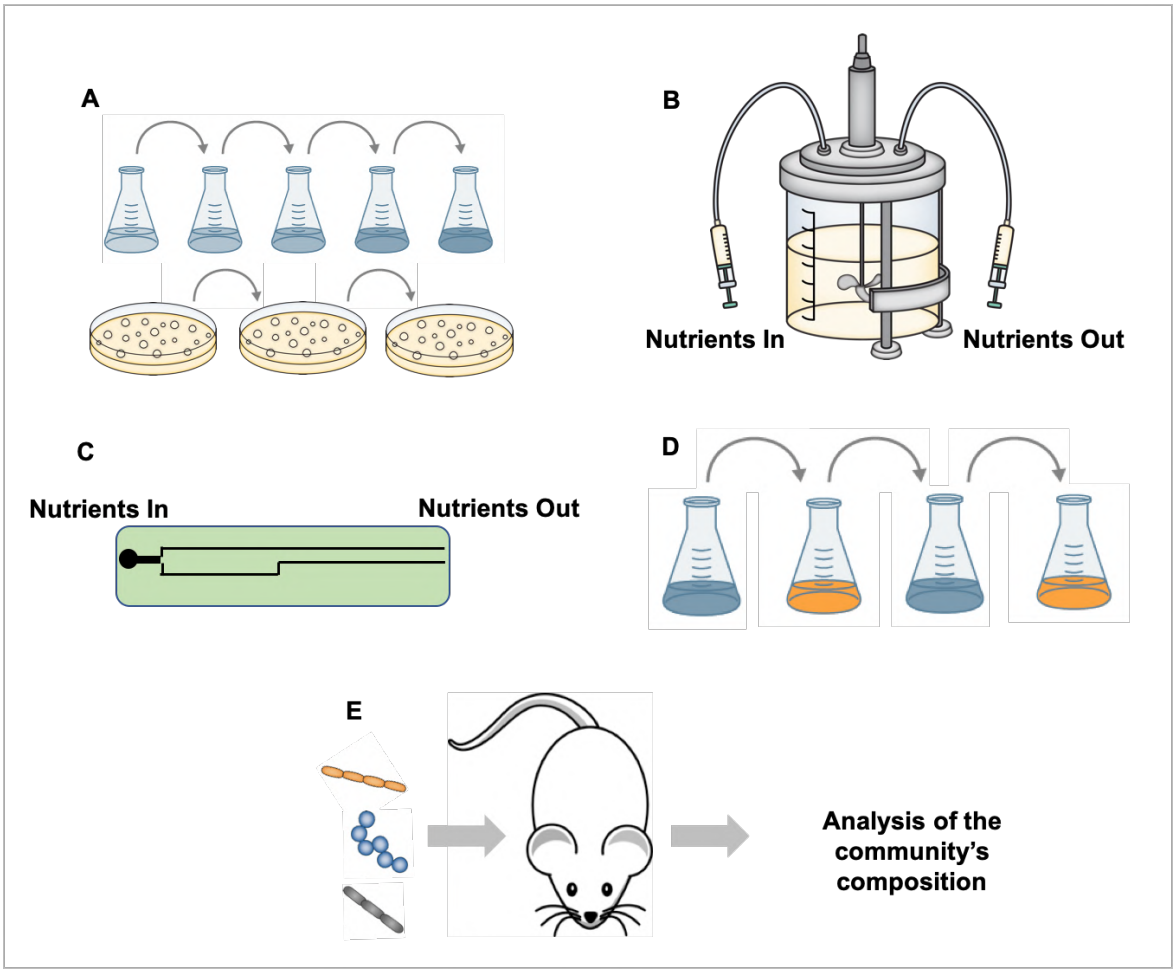

Figure 1.1. Experimental setup of ALE. (A) Batch culture adaptation can be performed either in liquid or solid media and can be easily scaled to a large number of replicates. (B) Chemostat cultures have a constant supply of fresh medium, keeping growth constant and the population size relatively stable. (C) Custom designed microfluidics systems are used for precise control of nutrient supplementation. (D) Alternating environmental conditions can be used to simulate dynamic environmental conditions and nutrient availability. (E) Microbial cultures can be propagated in a model organism, recovered and analysed, or propagated again.

1.4 The role of evolutionary biology in biotechnology

Microorganisms have been used by mankind throughout the centuries, with the most common examples being food fermentations that lead to the production of alcoholic beverages (wine and beer) and products that can be stored for longer periods (cheese and pickled vegetables). Since the 1980s, when it was first introduced, the term industrial biotechnology has become increasingly popular and is expanding rapidly. This term is being used to describe the use of living organisms (fermentation), or either their purified enzymes or metabolites (cell free catalysis), to produce valued chemicals, food products or energy at a large scale (Otero and Nielsen, 2010). Industrial use of microorganisms is increasingly drawing the attention of both scientists and the public, since the demand for specialised compounds through sustainable processes has increased, as well as the intention to decrease the use of fossil fuel, at least for the production of chemicals (Lee *et al*., 2012). A large arsenal of tools to manipulate DNA is available and is used to engineer the desired characteristics in the selected microbial strain, depending on the application. However, ALE has been established as a popular tool to either supplement or substitute the use of genetic engineering, and this way evolution successfully transitioned from ecology to biotechnology.

Most commonly, ALE is used complementary to metabolic engineering, in a term defined as evolutionary engineering. Evolutionary engineering is describing the screening process and selection of cells that express a desired phenotype and that they may originate from cell libraries, are created adaptively, through ALE, or randomly through random mutagenesis (Lee *et al.*, 2012). ALE can also be used to 'heal the scars' that are a by-product of the engineering process. These scars are usually off-target effects or unwanted characteristics that have been introduced into the genome of the microorganisms (Tenaillon, 2018). For example, methods based on recombination of DNA sites may lead to recombination of larger regions of the chromosomes. As a result, undesired point mutations are introduced into the genome, or regulatory elements may be shifted and therefore gene expression might be altered, facts that influence the phenotype and fitness of the cell. The use of ALE allows for the purge of cells with defective phenotypes. In addition, it has been observed that engineered strains exhibit decreased fitness when grown under the desired conditions for the selected application, due to the fact that the laboratory environment is fundamentally different from large-scale industrial conditions. In this case, repeated cycles of ALE lead to improved growth under industrial conditions, optimizing the phenotype (Figure 1.2C). Moreover, ALE can be used to evolve microorganisms to resist high concentrations of accumulated toxic by-

products (Dragosits and Mattanovich, 2013) and thereby towards improved yields and reduced costs (Portnoy *et al.*, 2011) in an industrial setting (Figure 1.2D).

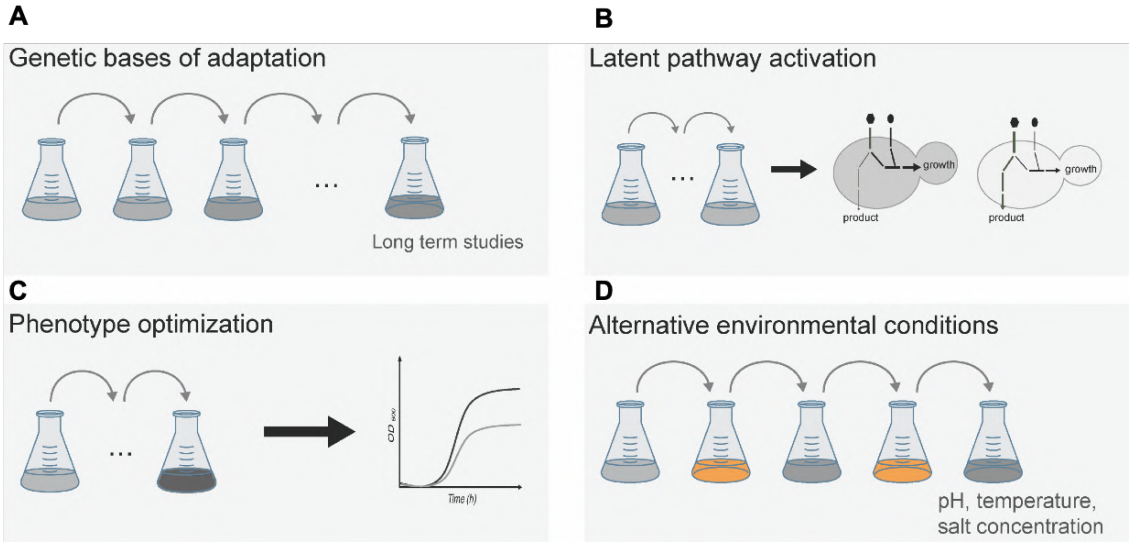

Figure 1.2. Research questions that take advantage of ALE. (A) Elucidation of genetic bases of adaptation that can be used as targets for future genetic engineering. (B) Improvement of complicated traits when engineering approaches are not straightforward. (C) Growth optimisation under specific conditions, like industrial settings. (D) Adaptation to harsh environmental conditions, leading to more robust strains and higher yields.

Metabolic engineering is defined as the directed modulation of metabolic pathways using methods of recombinant DNA, aiming to overproduce valuable compounds, such as fuels and pharmaceutical products from economic resources (Stephanopoulos, 2012). Metabolic engineering remains an extremely complicated process, without a certain outcome, as production yields are influenced by numerous parameters. For example, many of the catalytic steps of the carbon fluxes in a biosynthetic pathway may act as bottlenecks, limiting the yield of the end product (Dietrich *et al.*, 2010). Moreover, the genetic regulatory system of the cell and its interaction with the targeted pathway often affect the final yield of the desired compound. Therefore, rational engineering of metabolic pathways is not always the best solution, but the use of ALE allows scientists to understand the adaptation mechanism of microorganisms and to select appropriate targets for genetic engineering (Figure 1.2A). In some cases, it is necessary to force microorganisms to utilise non-preferred substrates, without prior knowledge of the necessary metabolic steps. The use of ALE for these cases has led to an improvement of enzymatic kinetics and differences in transcription that would not have been predicted *a priori* (Portnoy *et al.*, 2011), with different beneficial mutations arising in parallel across the whole genome (Dragosits and Mattanovich, 2013).

Finally, adaptive evolution has developed into a popular tool used in applications where due to legislation or concerns of the general public, the use of engineered organisms is restricted or even forbidden. For example, the European Union (EU) has applied a strict legal framework, allowing only the cultivation of genetically modified maize, as a precaution to prevent potential harm to the stability of the ecosystems and the health of humans and animals alike. These measures were adopted due to the concerns expressed by environmental activists, farmers and skeptical consumers. Until now, it is forbidden to place on the market either food containing/consisting of genetically modified organisms (GMO), or food products produced by GMOs (Reg. (EC) No. 1829/2003, supra note 3, Recital 16). Current genetic tools, such as CRISPR, can be efficiently used to improve strains for production in numerous applications. However, these engineered strains cannot be used for food, feed, and beverage applications, due to the current legal framework and the concern of the general public about the use of GMOs (Burgess *et al.*, 2006). ALE is a powerful tool to create improved strains that can be used for this kind of applications. Unfortunately, it has been difficult, if not impossible, to use adaptive evolution to select for traits that are not associated with cellular fitness.

1.5 Thesis motivation and objectives

Adaptive laboratory evolution has secured a spot among the extensively used tools for metabolic engineering, such as genetic manipulation, genome scale modeling, fermentation methods and omics technologies, but it is not a universal method that can be applied indiscriminately and without careful prior considerations. One important aspect that needs to be taken into consideration is the so called First Law of Directed Evolution, which states that “you get what you screen for” (Schmidt-Dannert and Arnold, 1999) and was used to describe processes concerning directed evolution of industrially important enzymes. Thus, the successful use of ALE is dependent on the existence of appropriate screening and selection tests that will allow the identification and isolation of the best performing evolved strains that arise from a mixed population. The First Law can be modified into “you get what you select for”, to describe the biggest bottleneck of using ALE. The successful outcome of an ALE experiment depends largely on the selection of the appropriate environment and in extent evolutionary pressure, that will allow natural selection to act and confer an advantage towards the desired phenotypes. During ALE of microorganisms, an initial strain is transferred for a long period of time under controlled environmental conditions until mutations arise and a mixed population of strains is formed. The cells that are transferred throughout the duration of the experiment are the ones that are able to survive and proliferate

in the environmental conditions. Thus, as it can be understood, ALE can be applied for the production of important compounds which are connected to the central cell metabolism and are necessary for survival, such as the long-term evolution of a *Leuconostoc mesenteroides* strain, aiming to increase its tolerance and production of lactic acid (Ju *et al.*, 2016). No prior knowledge of the mechanism is necessary, as long as the desired trait can be coupled with growth (Winkler et al., 2013).

However, numerous microorganisms have the ability to synthesise industrially important compounds through their secondary metabolism. These metabolites include food aroma compounds produced during fermentation, such as diacetyl, a compound with buttery aroma, which is produced by lactic acid bacteria or higher alcohols and esters that are formed during the alcoholic fermentation of grapes by *Saccharomyces cerevisiae*. Production of such chemicals for food applications through genetic engineering is still not an option due to legislation or concerns of the general public. Furthermore, the natural synthesis of such compounds is not always directly connected to the general fitness of the cell and it cannot be coupled with an environmental pressure in a straightforward manner, like the utilisation of a non-conventional carbon source, for example. This fact renders ALE an unrealistic strategy for the evolution of microbial strains that can be potentially used in an industrial setting. What is more, the evolution process leads to a mixed population and usually laborious screening processes are necessary for the evaluation and final selection of the best performing strains. As a result, laboratory evolution needs to be transformed from adaptive to directed, meaning that the evolution process will lead the majority of the cells to the desired phenotype, increasing the confidence of the whole process and decreasing the required screening time.

The aim of the present study is to propose two distinct methodologies for direct evolution of costly metabolic traits that do not confer a fitness benefit, under normal environmental conditions. The term fitness in this context is used to describe the ability of a microorganism to adequately grow under specific environmental conditions and nutrient availability, so it can go through a serial transfer experiment. In Chapter 2, a novel approach of using mutualistic microbial communities for the increased production of health promoting food compounds is described. For this purpose, different species of natural vitamin producing lactic acid bacteria were combined with engineered *S. cerevisiae* under appropriate conditions that allowed the exchange of metabolites between them, aiming to connect any potential improvement of the general fitness of the community with a subsequent improvement of the vitamin secretion. The original idea was that both members of the community will invest resources to improve the fitness of their partners and in return receive

higher amounts of necessary metabolites to then improve their fitness. Through this evolutionary procedure, a number of expected and unexpected questions arose, which are discussed and explained throughout the whole second chapter of this thesis. Lastly, a mechanism for the detected changes that led to the phenotypic differences is proposed, as well as the necessary conditions to achieve such mutualistic evolution are discussed.

In the second part (Chapter 3), the application of a novel computational algorithm, *EvolveX*, that combines genome scale metabolic models and flux balance analysis, is described. This algorithm computes the environmental conditions that will couple cell growth to increased metabolic flux through target pathways. The increased fluxes will result to the improved production of the target compound, when the evolved organisms are reintroduced in the native environmental conditions (target niche). The model-designed chemical environment (a set of nutrient components and inhibitors) is used for a laboratory evolution experiment with straightforward growth selection on the cells. When the evolved population is subsequently switched back to the target niche, where they are expected to manifest the enhancement in the desired trait. The validity of this formulation is showcased by evolving wine yeast strains with selective increase of aroma compounds originating from the branched-chain or the aromatic amino acid pathway. Moreover, the precision of the *EvolveX* algorithm is evaluated, examining whether the evolution process is truly guided and which were the mechanisms that led to the phenotypic differences between the original parental strains and the evolved isolated clones.

Chapter 2: Directed Coevolution for Improved Production of Nutraceuticals

2.1 Introduction

The majority of microorganisms live as part of complex communities and not in isolation. As a result, the research focus of microbiologists has shifted towards the study of these microbial communities. However, ALE has not been established yet as a popular tool for the evolution of microbial communities and to our knowledge it has never been used to improve biotechnologically important traits in the level of the community. Towards this direction, we propose that metabolic interactions between the members of a community can be used as a new method of laboratory evolution. In this introductory chapter, we present valuable information about the physiology and the biosynthetic capabilities of microorganisms that we chose as an experimental model. Moreover, we describe in detail the importance of our target metabolites, as well as the types of interactions that have been identified described previously in literature.

2.1.1 Lactic acid bacteria: an important bacterial clade for the food industry

Lactic acid bacteria (LAB) constitute a group of Gram-positive non-sporulating bacteria that belong to the phylum *Firmicutes* (Smid and Kleerebezem, 2014). They form an ecologically diverse group of microorganisms united by the formation of lactic acid as the primary metabolite as an outcome/end-product of sugar catabolism (Fugelsang and Edwards, 2007). They generally have small genomes ranging from 2 to 3 Mb and for many of them the whole genomic sequence is available (de Vos and Hugenholtz, 2004). LAB species are usually isolated from nutrient-rich environments such as dairy products, plant material and the gastrointestinal tract of humans and animals (Claesson *et al.*, 2007). Lactic acid bacteria are widely used in the food and health industry and are selected for their ability to tolerate stress conditions such as high ethanol levels and low pH values (Sauer *et al.*, 2017). Moreover, lactic acid bacteria can produce numerous chemicals and antimicrobial compounds, such as organic acids, hydrogen peroxide, diacetyl and bacteriocins (Guo et al., 2012). The long history of using this bacterial species with their broad spectrum of applications has led to the appointment of a generally recognised as safe (GRAS) status by both the American and the European health agencies. The ecological diversity of LAB and the different niches they occupy is reflected in their genomic content and their metabolic capabilities. Comparative genomics of different species has shown that some species express a higher number of proteins for the metabolism of carbohydrates, lipids and amino acids, while others depend on transporters and extracellular proteases in order to take advantage of environmental resources. Species that express multiple proteases, peptidases and transporters, lack the ability to synthesize *de novo* amino acids, nucleotides, vitamins and co-factors (Claesson *et al.*, 2007). Another trait that diverges between LAB species is connected

to their central carbon metabolism, more precisely, their ability to use the same compound either as an energy source or as an electron transporter.

Fermentation is established as one of the most common and acceptable methods to increase food safety, prolong the shelf life of products as well as improve its organoleptic attributes, by enriching the available nutrients (Widyastuti *et al.*, 2014), long before the discovery of microorganisms. Increased health benefits from the consumption of fermented milk products were hypothesised already at the beginning of the 20th century by Ilya Mechnikov, who associated the longevity of Bulgarian farmers with the increased consumption of yoghurt. Mechnikov was also convinced about the health promoting abilities of lactic acid bacteria. The ability of these microorganisms to confer health benefits upon consumption was identified with further research. Roughly at the same time as Mechnikov, the French paediatrician Henry Tissier observed the lack of specific bacteria in the stool of infants suffering from diarrhoea and in 1965 was introduced the term probiotic for the first time, aiming to describe a life promoting compound of microbial origin which would be the opposite of antibiotics (Butel, 2014). Nowadays, according to WHO, the term probiotic defines "live microorganisms which, when administered in adequate amounts, improve the health of their host' (WHO, 2001), as it was recently described in a randomised double-blind trial. In this study, a combination of *Lactobacillus plantarum* and prebiotics (fructooligosaccharide) had a protective effect on newborns in areas of rural India with high infant mortality rate, due to sepsis (Panigrahi *et al.*, 2017). The probiotic activity as well as the fact that milk is the natural ecosystem of lactic acid bacteria has ranked these species as the most popular milk fermenters and are used in the fermentation process of meat, cereals and vegetables, with the production of LAB starter cultures exceeding 1.5×10^6 tons per year (Teusink and Smid, 2006). However, their use is not limited only to the food industry, as these bacteria are used for a variety of applications, such as bulk production of lactic acid. Alternatively, by diverting the glycolytic part of their metabolism from lactate to pyruvate, they can turn into efficient producers of flavour and aroma compounds like diacetyl, acetaldehyde and D-alanine. Finally, different species of LAB have been successfully engineered for protein and exopolysaccharides production, have been developed as live vaccines for the delivery of therapeutic compounds to their hosts (Bravo and Landete, 2017, Goh and Barrangou, 2019), or used for the production of small antimicrobial peptides called bacteriocins. It is therefore obvious that the group of lactic acid bacteria plays an important role in human nutrition and well-being, and that their potential can be expanded.

2.1.2 Nutraceuticals and their significance

2.1.2.1 A novel term for the description of valuable food compounds

Over the past decade, a large number of food compounds have been released to the market labelled as nutraceuticals. The term nutraceutical, originating from the words “nutrition” and “pharmaceutical” is used to describe food or food components with a claimed medical or health benefit and is best known for supplying essential minerals or co-factors to the consumer (Hugenholtz and Smid, 2002). Nutraceuticals may be macronutrients, micronutrients, or non-nutritive compounds naturally found in food or added during the production process (LeBlanc *et* al., 2011). Major groups of nutraceuticals include vitamins, polyols (known also as low-calorie sugars), prebiotics and antioxidants (Hugenholtz and Smid, 2002) and all of them can diverge from the microbial metabolism. Food that contains such health promoting compounds is termed as *functional food*. The demand for these products is increasing along with the awareness of consumers on the importance of balanced nutrition, leading to a $50 billion market share in 2006 (Burgess *et al.*, 2006). In the meantime, malnutrition remains an important issue in the developing world and many people are suffering from dietary deficiencies connected to insufficient vitamin uptake (Burgess *et al.*, 2009).

2.1.2.2 Importance of vitamins for the human physiology

Vitamins are essential micronutrients that participate as precursors, or cofactors in numerous significant metabolic reactions. Most of the times, they act in synergy with intracellular coenzymes, aiming to regulate intracellular biochemical reactions and are generally divided into two distinct groups, fat-soluble vitamins, like vitamins A, D, E and K, that include vitamin C and the B complex vitamins. The former can act as cofactors, or participate as parts of cell membrane, whereas the latter act mostly as coenzymes or carriers of specific chemical groups and can be used for the biosynthesis of amino acids and nucleic acids (Burges *et al.*, 2009, Capozzi *et al.*, 2012, Papagianni, 2012). Humans are not able to synthesise the majority of vitamins and for that reason need to obtain them through their diet, as vitamins are readily available in different amounts and combinations in all kinds of food. However, vitamin deficiencies are still present not only in underdeveloped societies, but in most highly industrialised nations as well, due to malnutrition and unbalanced diet (LeBlanc *et al.*, 2011), or because these compounds can be easily removed or destroyed during food processing (Capozzi *et al.*, 2012). Subclinical deficiencies still persist. For example, riboflavin deficiency is observed in population groups whose diet lacks dairy products and

meat, as well as in groups with increased dietary requirements. One such case has been described for adolescent girls in the United Kingdom and Ireland (Burgess *et al.*, 2009). Pathological vitamin deficiencies may also be associated with the effects of alcohol or drugs consumption or with specific physiological states, such as pregnancy and breastfeeding (Russo, 2014).

For this reason, many countries have adopted a system of fortification, which means that specific vitamins and minerals are added to certain food before they are sold for consumption. Most commonly wheat flour is fortified with iron, folic acid, thiamine, riboflavin and niacin in an effort to reduce incidents of anaemia and neural tube transformation (LeBlanc *et al.*, 2011). However, concerns about fortification have been raised as to whether elevated levels of vitamin consumption might lead to or mask other pathologies. Such an example is the excess intake of folic acid, which may camouflage the haematological symptoms of vitamin B_{12} deficiency. This deficiency is common among people above 50 years of age, as they cannot absorb the vitamin efficiently. Moreover, fortification levels are estimated based on the average healthy adult and consequently children might be exposed to extremely high concentrations that not necessarily offer any health benefit. Another well-studied example is that foetuses are exposed to increased folate levels during pregnancy, due to the fact that women are prescribed folic acid supplements in combination with the consumption of fortified food. It has been shown that increased folic acid levels promote the selection of a methylenetetrahydrofolate polymorphism that has been associated with some debilitating diseases (LeBlanc *et al.*, 2011). Therefore, it is believed that an alternative to fortification with chemically synthesized vitamins is needed, preferably a method that would allow the production of food containing elevated levels of naturally occurring vitamins. The use of natural vitamin secreting microorganisms for the production of fermented food with elevated nutritional value can potentially offer a promising alternative to chemical fortification, since they will increase the vitamin levels of food without surpassing the level of daily recommended intake (RDI).

2.1.2.3 Vitamin production in microorganisms

Even though the majority of prokaryotes exhibit requirements for vitamins and need to take them up from the environment, species with biosynthetic capabilities for various vitamins have been observed and described (Burgess *et al.*, 2004, Capozzi *et al.*, 2011, Thakur *et al.*, 2016). It seems that the ability of bacterial strains to produce vitamins is much more common than it was previously believed and nowadays it is well known that the gut microbiota is an important supplier of various vitamins to the human organism, with vitamin

K and various B complex vitamins of microbial origin being secreted and absorbed in the colon (LeBlanc *et al.*, 2013). The capabilities of various microbes to synthesise vitamins have been vastly exploited in industrial settings, with the global vitamin B market increasing by more than 4% annually. Up until now, the production of vitamins is based on either chemical synthesis or biotechnology. Biotechnological processes are constantly gaining ground over the traditional organic synthesis, which requires the use of non-renewable chemicals and leads to the production of dangerous by-products. Apart from environmental advantages, since bioprocesses reduced the environmental footprint by 43%, biotechnological approaches favour economic development as well and as a result, the production of vitamins through fermentation increased from 5 to 75% between the years 1999 and 2012 (Acevedo-Rocha *et al.*, 2019). An important percentage of the total vitamin market is divided among the B vitamin group, with the largest volume being used by feed, food, pharmaceutical and chemical industries. Most of these vitamins are synthesised chemically, however, there are established methods for biotechnological production of riboflavin and fortification of food products with riboflavin and folate. Initially production of riboflavin was based on naturally overproducing isolates, but these stains produce relatively small amounts and for that reason have been completely replaced by genetically engineered strains of *Bacillus subtilis* and *Ashbya gossypii*. These microorganisms have been engineered with the aim to impair regulation, increase the copy number of riboflavin biosynthesis genes or to increase the supply of the pathway with the necessary precursors, leading to yields up to 14 g/L. Riboflavin overproducing strains of LAB mutants have been isolated based on their ability to grow in the presence of roseoflavin, a toxic analogue of riboflavin. Growth in the presence of toxic analogues is a method of random mutagenesis, but the riboflavin yield of these strains is very low (0.5 to 0.6 mg/L for *L. plantarum* and 1.2 mg/L for *Lactococcus lactis*) compared to well-established industrial GMO producers. The big advantage of these natural isolates is the fact that they can be used as starter cultures in fermentation, aiming to increase the vitamin levels of food products. On the other hand, microbial production of folate is not considered economical enough yet, so microbial applications are limited to the fortification of dairy products (Abbas and Sibirny, 2011).

2.1.3 Riboflavin, a vital precursor

2.1.3.1 General characteristics of riboflavin

Riboflavin or vitamin B2 was first discovered in 1879 as a yellow pigment in milk and so was named lactoflavin. It was proved to be essential for growth and survival by Richard

Kuhn at the University of Heidelberg (Abbas and Sibirny, 2011). Riboflavin is a water-soluble vitamin with no direct metabolic function in living cells, it is nevertheless the precursor of two important electron carriers, Flavin mononucleotide (FMN) and Flavin adenine dinucleotide (FAD) (Figure 2.1). FMN and FAD participate in countless oxidation-reduction reactions and function as coenzymes for numerous enzymes, known as flavoproteins (Capozzi *et al*., 2012). Flavoproteins are so abundant that it is estimated that 1 to 3% of the genes in both eukaryotic and prokaryotic genomes encode such proteins (Abbas and Sibirny, 2011). Riboflavin is sensitive to light exposure and photolysis leads to the production of two compounds, namely lumiflavin and lumichrome, but these compounds lack specific metabolic importance. Riboflavin is synthesised by plants and microorganisms, but humans need to acquire it through their diet and the daily required intake (RDI) varies between 0.9 and 1.6 mg, depending on gender and age. Symptoms of ariboflavinosis, the clinical manifestation of riboflavin deficiency, include sore throat, hyperaemia, oedema of oral and mucous membranes, cheilosis and glossitis (LeBlanc *et al.*, 2011). Riboflavin is present in a large variety of food products, such as cereals, meat, fatty fish and green leafy vegetables, but milk and dairy products remain the main sources of riboflavin in Western societies (Burgess *et al.*, 2009). However, milk is not considered a good source for riboflavin uptake because it contains 1.2 mg/L and an adult would need to consume more than a litre of milk every day, in order to meet the RDI. In reality, the average consumption of milk is at least 5 times lower at around 200 mL per day (LeBlanc *et al.*, 2011). Research has shown that riboflavin concentrations vary between different dairy products due to the action of the microbial strains used during the fermentation process. For example, riboflavin content in buttermilk and yoghurt can reach up to 2 mg/L, a fact that highlights the importance of creating proper starter cultures with strains that can improve the nutritional value of food. Apart from being a dietary supplement, riboflavin is used as a food colouring agent due to its bright yellow colour, or as a treatment for migraines, malaria, or side effects of AIDs treatment (Burgess *et al.*, 2004).

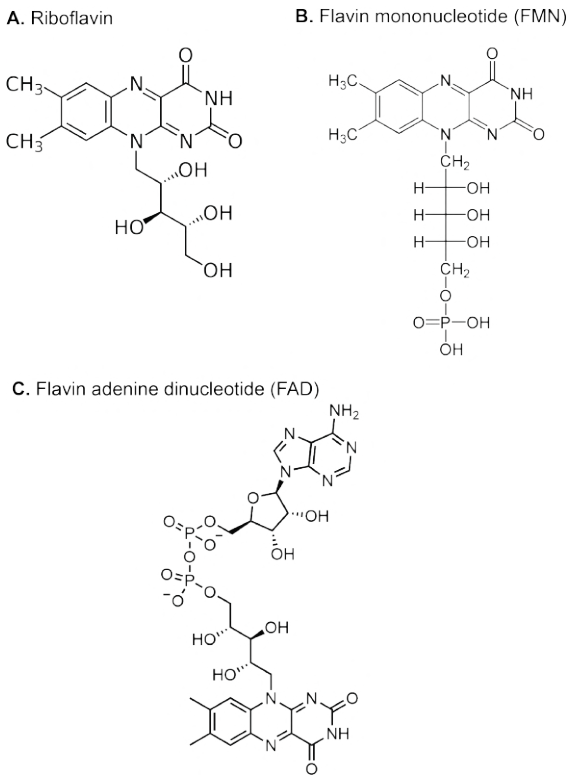

Figure 2.1 Chemical structures of the most important flavins. (A) Riboflavin, (B) Flavin mononucleotide (FMN) and (C) Flavin adenine dinucleotide (FAD).

2.1.3.2 Microbial production of riboflavin

Riboflavin is synthesised by numerous bacteria and the biosynthetic pathway has been described for both Gram-positive and Gram-negative bacteria with the biggest emphasis on *Escherichia coli* and *B. subtilis* (Thakur *et al.*, 2016, Figure 2.2A). Briefly, for the initial step of the biosynthetic pathway, the precursors GTP and ribulose 5-phospate are needed. First of all, the imidazole ring of GTP is hydrolysed with subsequent release of formate and pyrophosphate in a reaction catalysed by the enzyme GTP cyclohydrolase II, encoded by the gene *ribA*. The product of this first reaction, 2,5-diamino-6-ribosylamino-4(3H)-pyrimidinone 5'-phosphate, undergoes a two-step reaction, to be converted to 5-amino-6-ribitylamino-2,4(1H,3H)-pyrimidinedione 5'-phosphate. The two-step reaction includes the hydrolytic cleavage of the amino group at position 2 of the heterocyclic ring and the reduction of the ribosyl side chain by a NADPH depended reductase, thus providing the ribityl side chain of the vitamin. These two reactions occur in different order in different microorganisms, with the deamination occurring first in bacteria, while in yeast and fungi the reduction is the first one to take place. Before any further conversion, a dephosphorylation step to remove the 5-phospate tail is necessary, but the exact reaction step remains undescribed. The next step involves condensation of the dephosphorylated pyrimidinedione with 3,4-dihydroxybutanone 4-phosphate by an enzyme called 6,7-dimethyl-8-

ribityllumazine synthase (RibH) and in the final step of the pathway, the produced 6,7-dimethyl-8-ribityllumazine is dismutated by the enzyme riboflavin synthase (RibB). This reaction leads to the formation of a second compound which can be recycled in the biosynthetic pathway. In conclusion, the formation of one riboflavin molecule requires one molecule of GTP and 2 molecules of ribulose 5-phosphate, meaning that out of the 17 carbon atoms found in the riboflavin molecule 13 derive from the pentose phosphate pool (Bacher *et al.*, 2000). As it was mentioned already, riboflavin is the precursor for FMN and FAD and a specific phosphorylation of the ribityl chain catalysed by the riboflavin kinase leads to the formation of FMN, while FMN is turned into FAD by the enzyme FAD synthetase, also known as FMN adenyltransferase, as it catalyses the transfer of adenylyl moieties from ATP to FMN. Conversion of FAD to FMN and AMP is achieved by a group of enzymes called FAD pyrophospatases, while FMN can be degraded again to riboflavin through the action of a nonspecific phosphohydrolase (Abbas and Sibirny, 2011).

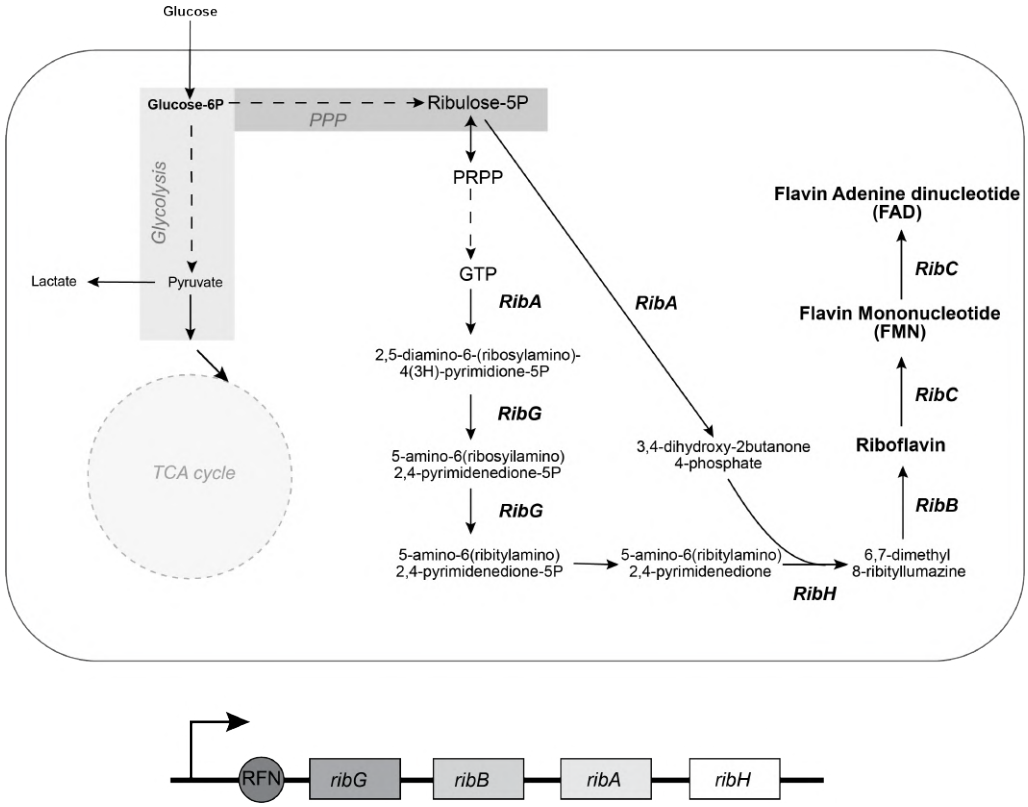

Figure 2.2 Biosynthesis of riboflavin in bacteria. (A) Biosynthetic pathway of riboflavin in bacteria. Full arrows represent reactions that happen in one step, while dashed arrows represent several intervening reactions. (B) Organisation of the riboflavin operon in bacteria. Each gene is depicted as a rectangle, while the regulating riboswitch is visualised in a circle.

In bacteria, the genes that code for all the necessary proteins that are involved in the biosynthetic pathway of riboflavin form an operon, known as the *rib* operon. This operon is made up by 5 separate genes (*ribGBAHT)* and is transcribed as one polycistronic RNA with

a total length of 4.3 kbp. There is also an untranslated region of about 300 bp, termed *ribO*, located upstream of the first gene, which was believed to be the operator region. Regulation of the operon's transcription is mediated through a riboswitch (RFN element), this small sequence of 140 bp is located in the leader region of the *rib* operon and is able to fold into an evolutionary conserved RNA structure. This structure can bind directly with FMN, forming a terminator hairpin, bringing the transcription of the operon to a halt (Abbas and Sibirny, 2011). It is worth mentioning that the riboflavin producing phenotype is strictly connected with the structural integrity of the operon and is often a strain specific ability, as it has been described that multiple strains lack parts of genes *ribG* (also termed as *ribD*) and *ribB* and hence lack the biosynthetic ability. One such example is the *L. plantarum* strain WCFS1, which, while being able to produce and secrete folate, cannot grow efficiently in the absence of riboflavin. This strain is widely used and its completely assembled and annotated genome remains the reference genome for the species *L. plantarum* (Capozzi *et al*., 2012, Figure 2.2B). Lactic acid bacteria that have been exposed to either toxic analogues of purines or roseoflavin have been used by researchers to create fermented milk products with elevated concentrations of riboflavin and consumption of these products improved the phenotype of riboflavin deficient lab animals (Burges *et al.*, 2006).

2.1.4 The B9 vitamin family

2.1.4.1 Folic acid and its vitamers

Another important member of the B group vitamins is folate and its bioactive forms (vitamers), known also as the B9 vitamin family, which includes naturally occurring polyglutamates in food as well as folic acid which is a synthetic folate used for food fortification and in dietary supplements (Meucci *et al.*, 2018). Folate consists of a pteridine moiety connected to para-aminobenzoic acid (PABA) through a methylene molecule and PABA itself being connected to one or more molecules of L-glutamic acid (Figure 2.3, Saubade *et al.*, 2017). Folate is actively taking part in numerous essential cell functions, ranging from amino acid metabolism, DNA replication, repair and methylation to the biosynthesis of other vitamins and thus is necessary for cell division (Laiño *et al.*, 2013, Saubade *et al.*, 2017). As in the case of riboflavin, humans are unable to synthesise folate and need to acquire it through nutrition with folate deficiency being connected to a variety of physiological disorders, such as megaloblastic anemia and neural tube defects in infants (Wegkamp *et al.*, 2007). In adults, folate deficiency has been associated with an elevated risk for colon cancer and cardiovascular diseases, as well as with cognitive disorders like

Alzheimer's and even with depression (Iyer and Tomar, 2009). Even though folate can be taken up by the consumption of legumes, vegetables, liver, milk and other fermented dairy products, daily intake might still remain below the required levels which range between 200 and 400 µg for adults and therefore folate deficiency remains a problem in both developed and developing countries. For this reason, folic acid is used in industry for food fortification, however, numerous studies question the safety of consuming chemically synthesised folate and suggest natural fortification through microorganisms (Meucci *et al.*, 2017).

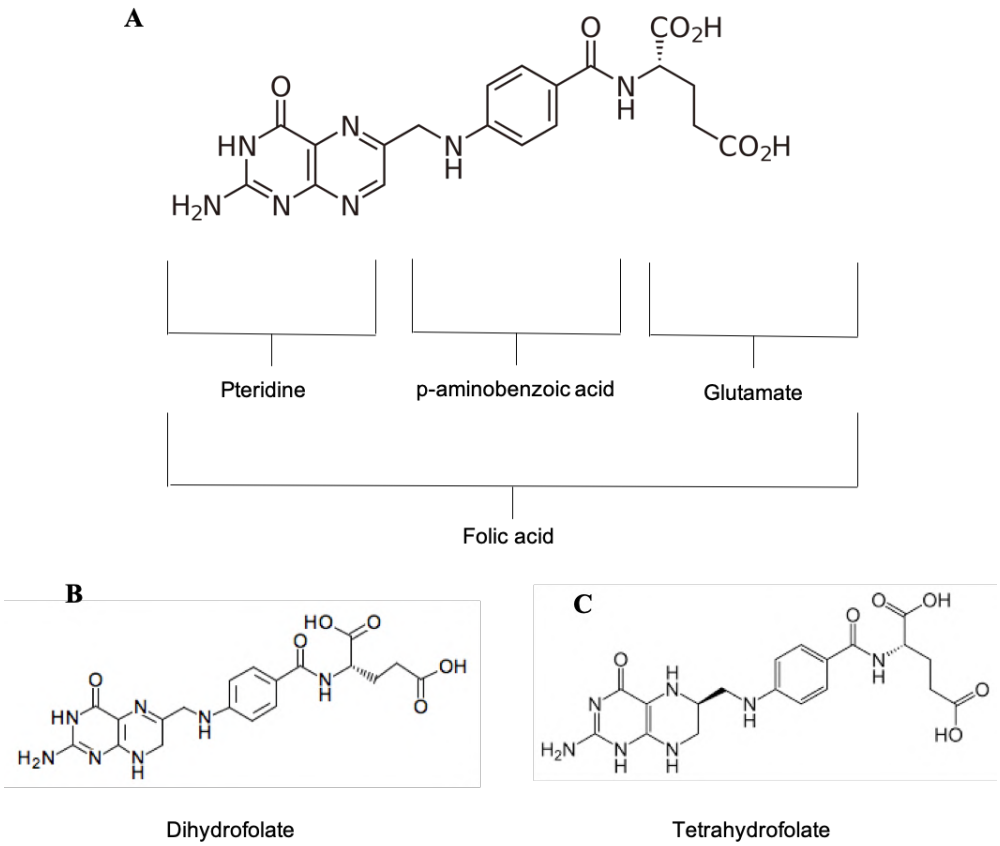

Figure 2.3. Chemical structure of the B9 family vitamins. Chemical structure of (A) folic acid and its active forms (B) dihydrofolate (DHF) and (C) tetrahydrofolate (THF).

2.1.4.2 Folic acid production in bacteria

The ability of bacteria to synthesise folate has been identified in different species and has been described extensively in the literature. The folate biosynthetic pathway consists of the reactions needed to produce the three distinct parts (pterin, PABA and glutamate) that combine to form folate. In the first steps, GTP is transformed into the active cofactor tetrahydrofolate (THF) through two distinct reactions. In the first one, PABA is condensed with 2-amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine to produce dihydropteroate and in the second one glutamate is connected to dihydropteroate leading to the formation of dihydrofolate. The necessary precursor PABA can be synthesised in bacteria from D-erythrose 4-phospate and phosphoenolpyruvate through the pentose phosphate pathway, leading to the final formation of chorismate. Most of the times, glutamate is taken up from the environment (Iyer and Tomar, 2009, Wegkamp *et al.*, 2007). In total, the production of folate requires 17 enzymatic reactions, which can be extended to 45 with the synthesis of the rest of the vitamers (Saubade *et al.*, 2017) and the folate gene cluster consists of 5 genes with an approximate size of 10 kb (Iyer and Tomar, 2009). In *Firmicutes*, a riboswitch able to bind to THF was recently discovered and it is believed that it associates both with mRNAs that encode genes responsible for the transport of folate and biosynthetic genes like *folC* and *folE* (Trausch *et al*., 2011).

In contrast to the case of riboflavin, where scientists focus on the use of toxic analogues to isolate overproducing mutants, overproduction of folate by LAB has been achieved through methods of genetic engineering. One such case is the species *Lactobacillus gasseri*, which under normal conditions consumes folate, but after the transfer of a plasmid that contained the folate gene cluster from *L. lactis* it transformed into a folate producer. Furthermore, *L. lactis* strains were engineered to retain folate inside the cell and when they were administrated to lab animals lacking folate in their diet. Folate levels were normal in the organ and the blood samples of those animals a fact that indicates lysis of the bacterial cells in the intestinal tract and uptake of the released folate (Iyer and Tomar, 2009). Even though increased production of folate can prove useful in a variety of applications, transfer to an industrial scale would be challenging, since engineered overproducing strains often exhibit reduced growth rates (LeBlanc *et al.*, 2011). However, consumption of genetically engineered microorganisms by humans is still not allowed, meaning that different strategies need to be followed for the improvement of strains that would be used for the creation of food products with increased nutritional value.

2.1.5 The social life of microorganisms

2.1.5.1 From single species studies towards microbial communities

In the early history of microbiology, scientists like Robert Koch and Louis Pasteur, focused their research on the isolation, cultivation and study of single microbial species. These types of studies proved extremely useful for the understanding of microbial physiology and biochemistry. However, it is now understood that in reality microorganisms live in complex environments, co-existing with countless other species in structured communities. The term community is being used to describe the sum of species, or consortia that coexist at the same place at the same time (Boon *et al.*, 2014). Such microbial consortia affect directly biogeochemical cycles and play a role in human physiology (McCully *et al.*, 2017) and hence scientific focus has shifted towards the elucidation of the interactions that occur within these communities.

2.1.5.2 Forms of interaction in microbial communities

So far, a large number of different forms of interaction have been described, but in general they fall into three main categories, as they can be antagonistic, beneficial, or neutral depending on whether the outcome of the interaction is negative, positive or there is no visible interaction. The type of interaction may vary even more, as it may be either obligate or facultative (Boon *et al.*, 2014). A well-studied example of antagonistic interaction is the secretion of compounds, like toxins, antibiotics, or lactic acid that aim to offer a growth advantage to one species by inhibiting the growth of other microbes in close proximity (D'Souza *et al.*, 2018). Other examples of negative interactions include predatory and parasitic, or cheating, behaviour (Hoek *et al.*, 2016). Cheating behaviour typically arises in resource limited environments when the survival of the community is dependent on the exchange of metabolites, but a proportion of cells only benefits from these common goods without contributing its share. According to evolutionary theory, natural selection should favour such selfish behaviours that increase the fitness of an organism at the expense of the fitness of the other community members (D'Souza *et al.*, 2018). However, cooperation seems to be of equal importance in microbial communities, since the products of the metabolism of one strain may be used as nutrition by other members of the community. This phenomenon, known as cross-feeding (Smith *et al.*, 2019), is met abundantly in nature under different aspects.

Nutrient exchange in bacteria has many different faces and in literature different terms have often been used to describe the same phenomena. Classical definitions were focused on

syntrophy, meaning the obligatory exchange of necessary metabolites between two species (Pande and Kost, 2017). This definition is excluding cases of positive interactions that occur through detoxification, more specifically the case where one strain is metabolizing one compound that is toxic for another species, thus allowing the second species' survival. Moreover, it has been described in literature (Hillesland, 2018) the idea that these mutualistic dependencies are not fixed and can be used for any two-way exchange of nutrients. The lack of a clear definition led Smith *et al.*, to propose in their recent review (Smith *et al.*, 2019), the term metabolic cross-feeding to include all possible forms of positive interactions under one definition. Therefore, metabolic cross-feeding is used to describe an interaction that occurs between microorganisms when an end point product of the metabolism of one species is metabolised further by another species and it can be either a one way or a two-way interaction (Figure 2.4).

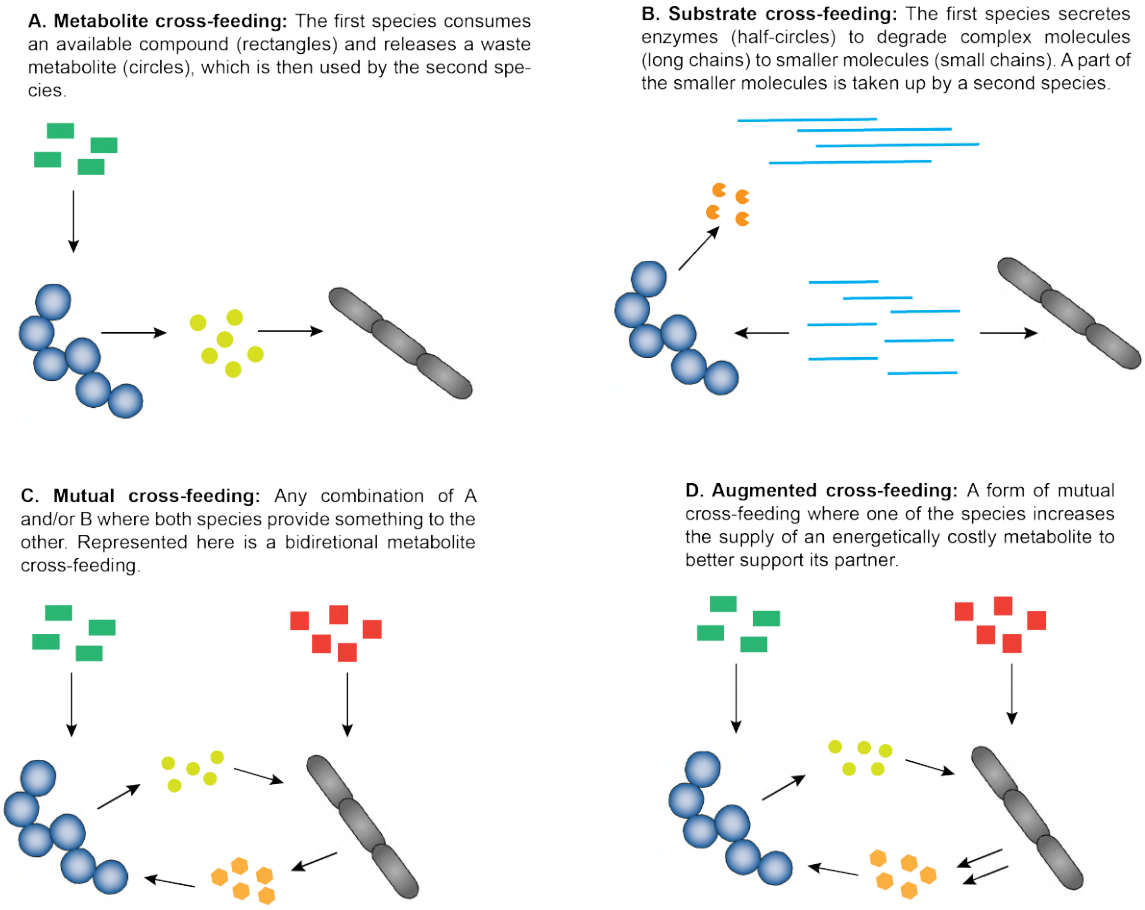

Figure 2.4. Examples of microbial metabolic cross-feeding. (A) In classic metabolic cross-feeding, the secreted by-products of one species are taken up and metabolised by a second one. (B) Alternatively, one species is enzymatically breaking down substrates which are partially available to other species as well. (C) In mutual cross-feeding, two species secrete compounds that are useful to the other species, while this exchange can be actively increased (D) to better support one of the partners (adapted from Smith *et al.*, 2019).

2.1.5.3 Metabolic cross-feeding: single direction interactions

There are available experimental data that describe different forms of metabolic cross-feeding. The first class of such-cross feeding are cases of one way interactions where one species is releasing by-products into the environment, also known as “public goods”, which in turn are taken up by another species that further metabolises them for energy production. Such a case has been described in Lenski’s long-term evolution experiment where *E. coli* mutants which preferably utilized acetate arose in the population and were able to consume this unexploited carbon source (Lenski, 2017). Similar to the release of by products is the release of metabolites due to leakage or overflow of metabolism, with such a case having been described for yeast secreting amino acids in the environment and subsequently allow the growth of lactic acid bacteria which exhibit multiple amino acid auxotrophies (Ponomarova *et al.*, 2017). Apart from the release of ready to use metabolites, it is also known that some microorganisms release enzymes to degrade complex compounds, so they can import them in the cell environment and metabolize them, but the broken-down compounds are available to other species as well. Such a case has been described in co-cultures human gut species where *Bacteroides thetaiotaomicron* is able to degrade extracellularly oligofructose and inulin, turning them into short chain sugars. The other bifidobacteria in the co-culture lack the particular ability, but are able to uptake the short chain sugars from the environment. However, the time period between the enzymatic breakdown and the uptake is favouring competition between the enzyme producer and the rest of the close by species and as a result enzyme secretors often show reduced fitness when are grown in co-cultures (Smith *et al.*, 2019).

2.1.5.4 Metabolic cross feeding: bidirectional interactions

Bidirectional cases of metabolic cross-feeding can also be encountered. Such cases are usually defined as reciprocal, since both partners’ survival is connected to the exchange of products of the metabolism, which often may be energetically costly to produce. In natural ecosystems, cases of bidirectional exchange of by-products, termed also as proto-cooperation, have been described between ammonia oxidizing microbes and nitrite oxidizing bacteria, where the ammonia oxidizing microbes release nitrite which in turn is used by the nitrite oxidizing bacteria. The reciprocity between the species is attributed to the fact that nitrite oxidisers can convert urea to ammonia and carbon dioxide, thus supplying ammonia oxidisers with the necessary energy source (D’Souza *et al.*, 2018). On the other hand, so far it has proven difficult to describe the naturally occurring phenomenon of bidirectional metabolic cross-feeding when the exchange products are costly to produce. For that reason,

the best described cases originate from synthetically interactions that were engineered in the lab. Such an example was described by Mee *et al.* (2014) with the engineering of multiple *E. coli* strains, each one with an auxotrophy for an essential amino acid. The engineered strains were not able to grow in monoculture, but they could rescue each other's growth through the secretion of amino acids into the extracellular environment. In a similar fashion, Pande *et al.* (2014) engineered amino acid auxotrophies in *E. coli*, as well as mutations that added an overproducing phenotype for other amino acids. The researches in this study concluded that co cultures of auxotrophs exhibit higher growth rates compared to the prototrophic parental strain, as the cost of overproducing a certain amino acid was less than the benefit of receiving ready-made amino acids by their partner.

Engineered exchange of amino acids has become a popular tool among researches, since apart from the two previously described examples, two more studies utilise such dependencies to demonstrate how microbial mutualism may persist and evolve. In the first of these studies (Ziesack *et al.*, 2019), the researchers engineered amino acid auxotrophies to four different gut bacterial species and concluded that metabolically dependent communities were more stable and were exhibiting higher diversity even when they were introduced to the gut of gnotobiotic mice. Finally, an interesting observation has been made in co-cultures of *E. coli* and *Salmonella typhimurium* (Harcombe, 2010, Harcombe *et al.*, 2018). In the older study, the two bacteria are grown on solid media that contain lactose as a carbon source and therefore *S. typhimurium* was depended on *E. coli* to break down lactose and release simpler carbon sources, while *E. coli* was depended on the secretion of methionine by *S. typhimurium*. Strains of *S. typhimurium*, which were resistant to the toxic analogue ethionine were able to support the growth of *E. coli* and when the two species were grown in co-cultures, firstly for 3 days and then regrown for 5 days, *Salmonella* isolates secreted in the medium 15-fold more of the energetically costly amino acid. In a similar manner, *E. coli* and *S. enterica* were evolved together on lactose minimal media plates (Harcombe *et al.*, 2018) and by the end of the experiment, sugar secreting *E. coli* mutants were isolated. These mutants were secreting more energetically costly synthesised sugars, not just metabolic by-products, and as a result they exhibited worse fitness than the parental strain. However, the same mutants were more fit than their parental strain when grown in co-culture conditions. In conclusion, co-operation requires reciprocation, as it seems that an adaptation strategy where the individual costs come second to the fitness of the whole community is much more common than it was believed in the past. The question how such interactions may have evolved in the first place still remains an intriguing puzzle for many researchers, as it is not

easy to explain based on natural selection. Why would investing resources to benefit other non-relative species under specific environmental conditions be favoured by natural selection? Why on the other hand selfish strategies that would allow them to take advantage of their partner without keeping their part of the bargain are not selected for in nature?

2.1.5.5 The evolutionary mechanism of syntrophy

It is generally believed that syntrophy is an important tool of microbial communities in their effort to survive in constantly fluctuating environments, under conditions where each individual community member would in principle die out. Cooperation is mainly explained by Hamilton's rule, which states that the most important factor for selection to favour collaboration is that the benefits for the recipient times the relevancy of the provider and the recipient, outweigh the costs for the provider (D'Souza *et al.*, 2018). The most popular theory for the emergence of mutualistic interactions suggests that mutualism might start off as metabolite leakage or by-product secretion from one species and the secreted metabolites benefit a second species living nearby. In turn, the second species adjusts its metabolism to the new nutrient sources, adjusting the metabolic fluxes and can potentially start secreting a metabolite that could be used by another species (D'Souza *et al.*, 2018). Over the passage of time, these dependencies are established between the species and evolve into cross-feeding. Alternatively, metabolic dependencies may initially arise through random mutagenesis that leads to loss of gene function and disruption of the metabolic pathway, rendering a species auxotrophic. This type of reductive evolution is also known as the Black Queen hypothesis, which states that under specific environmental conditions it may be favourable for a species to lose the ability to perform a specific biosynthetic function. Gene loss seems a beneficial adaptive strategy, when a costly to produce compound is free in the environment, or is secreted by another member of the community. Moreover, it is hypothesised that selection sometimes favours smaller cell complexity and consequently smaller genomes, since simpler cells can focus their energy spending on growth and not towards the synthesis of complex compounds (Smith *et al.*, 2019). Research on engineered auxotrophs has showcased how fitness may improve through reductive evolution, but the fitness advantage was related to the availability of the required compound, as well as to which gene of the biosynthetic pathway was deleted (D'Souza *et al.*, 2014). It is also believed that metabolic dependencies may start spontaneously from random combinations of microbes that exhibit complementary biochemical capabilities (Libby *et al.*, 2019). Finally, cross-feeding relationships may potentially form due to phenotypic plasticity and not necessarily loss of function mutations. For example, *E. coli* is able to utilise both glucose and acetate as a carbon source, depending

on the environmental conditions. Therefore, some cases of metabolic dependencies may actually be the most energetically efficient strategy for a cell to survive under specific conditions and nutrient availability (Smith *et al.*, 2019).

2.1.5.6 Microbial communities as biotechnological tools

Apart from ecological studies, microbial communities are of great importance to biotechnology as well, as they offer an important alternative for the production of compounds that would not be easily produced by one single strain. The use of microbial communities is not a novel idea, since humans have been using communities in fermentations for thousands of years already. However, the design of robust and reproducible communities for consistent results is a more recent idea. The engineering of microbial communities has been characterised as synthetic ecology and its aim is the use of different cells as building blocks for the rational creation of complex systems (Pandhal and Noirel, 2014). The biggest advantage of using microbial communities is the observed division of labour among their members, which can be defined as the combined use of various specialists towards a collective objective. Specifically, metabolic processes are divided between different species and as a result from an initial substrate and through different intermediates the production of the desired compound is feasible. The most important advantage of division of labour is that negative feedback loops from the accumulation of intermediates and competitive interactions of enzymes inside of one cell can be avoided (Lindermann *et al.*, 2016). Moreover, this approach allows for the utilisation of complex substrates that are normally not metabolised by biotechnological workhorses like *S. cerevisiae* and *E. coli*, but other members of the community have the ability to turn these substrates into simpler ones that can be easily taken up and used. In addition, the presence of a specialist that can remove inhibitory intermediates from the environment leads to higher conversation rates of the substrate and to the accumulation of more biomass (Lindermann *et al.*, 2016). Another advantage of using microbial communities is the flexibility to efficiently catalyse processes through a variety of enzymes and pathways, which cannot be achieved with the use of a single species, due to the fact that the number of properties that can be engineered is limited before genetic instability arises (Jagmann and Philipp, 2014). Finally, communities behave more robustly than monocultures during environmental perturbations, which might not be important under the well-controlled conditions of a bioreactor, but it is an important parameter in open field applications, such as the bioremediation of toxic waste (Lindermann *et al.*, 2016).

Microbial communities are widely used for a variety of applications, in an effort to utilise the above-mentioned advantages they exhibit in industrial settings. These applications range

from the production of biofuels and the conversion of xylan to ethanol (Pandhal and Noirel, 2014), to food fermentation, such as milk fermentation to produce dairy products and the fermentation of cocoa beans, to produce cacao. Especially in the case of food fermentations, naturally occurring communities are preferred, since they have naturally evolved to alter the properties and nutritional value of raw food materials. As a result, there is no need to alter the physiology of the strains and it is easy to create different combinations for different end products. Such a case is the production of cheese where mixing different LAB species leads to the development of a specific flavour, aroma and texture profile (Ghosh *et al.*, 2019). Especially in the case of cheese production, interactions between the LAB starter cultures and different yeasts or moulds are important either because yeasts are supporting the growth of the bacteria, or because they lead to the production of aromatic compounds during the ripening process (Lourens-Hattingh and Viljoen, 2001).

2.2 Results and Discussion

Despite its popularity, to the best of our knowledge, ALE has not yet been extended to use microbial consortia to improve a biotechnologically relevant trait, or to improve the biotechnological processes that are performed by these consortia. The hindrance in expanding towards such applications can be linked to the main disadvantage of using ALE, which is the limitation to select traits that are directly linked to cellular fitness. Phenotypes with potential value in biotechnological applications are difficult to improve through ALE as well, even though they are directly linked with cell fitness. For example, strains that naturally produce and secrete vitamins can potentially prove useful in food applications, but the improvement of their producing capabilities through ALE seems complicated, as they already produce enough to cover their needs. Therefore, simply selecting them based on their ability to grow in the absence of these vitamins should not necessarily lead to improved production.

In this thesis, we propose that metabolic cross feeding can create the appropriate selection pressure for increased production and hence can be used as a new method of laboratory evolution. Specifically, we hypothesise that in a mutualistic community, where each species is auxotroph for a specific resource that is lacking from the environment, but is secreted by its partner, increased production of a costly compound required by the second species will result in higher rewards for the first one. As a result, increased production of these compounds will lead to improvement of the overall community fitness through a serial transfer experiment. We decided to test our hypothesis by improving the production capabilities of natural B group vitamin producing strains and particularly vitamins B2 (riboflavin) and B9 (folate), using small mutualistic microbial communities which consisted of two species.

2.2.1 ALE for increased production of riboflavin

2.2.1.1 Establishing a model community for coevolution

The first step of our experimental setup was to choose a pair of microbial species that would form a mutualistic community, with a nutraceutical molecule being one of the exchanged compounds. In a previous work in our group, a chemically defined medium (CDM35) was designed, which allowed *S. cerevisiae* to nutritionally support two species of lactic acid bacteria, viz., *L. plantarum* and *L. lactis*. In this medium, *S. cerevisiae* was shown to secrete, through nitrogen overflow, a set of amino acids that the two LAB species are auxotrophic for (Ponomarova *et al.*, 2017). Therefore, we chose CDM35 as the evolution niche for developing a mutualistic community. Production of riboflavin (vitamin B2) was

chosen as a target molecule, since it is known from the literature that many LAB strains can secret it to the environment. Therefore, we removed riboflavin from the medium, which was then named CDM34R-. The following step included testing the ability of bacterial strains to grow in CDM34R-, without supplementation of riboflavin and amino acids. The strains we tested originated from our group's collection of natural isolates. One strain of *Lactobacillus plantarum* (GCK strain) that was able to grow without riboflavin supplementation, but not without the addition of amino acids, was selected as a potential candidate.

The second member of the community was the yeast *S. cerevisiae* and specifically a laboratory strain (BY background) of which we knew that it can adequately support the growth of LAB. However, *S. cerevisiae* is prototrophic for riboflavin and therefore we had to engineer the auxotrophic phenotype, by deleting 2 genes in the riboflavin pathway. The double mutant strain (*Δrib4:rib5)* exhibited an autotrophic phenotype, since its growth was strongly diminished without the addition of riboflavin even in rich medium. Moreover, the double mutant *S. cerevisiae* retained the ability to support the growth of *L. plantarum* in CDM35. Similarly, when we cultured both strains in CDM35R-, we observed growth, meaning that they could interact with each other. Since the established community was able to grow in the CDM35R-, we were finally able to begin the directed evolution experiment (Figure 2.5A).

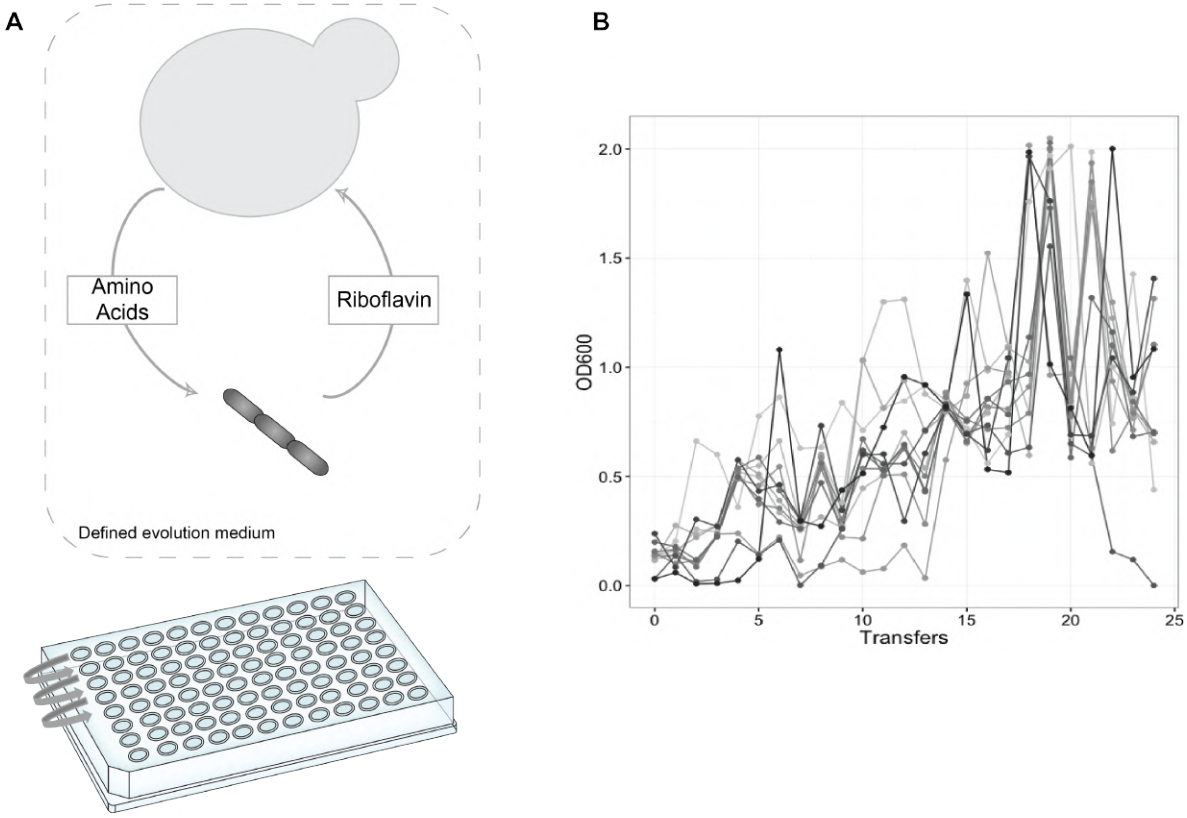

Figure 2.5. Experimental coevolution of mutualistic community for increased production of riboflavin. (A) Experimental design of co-culture evolution for increased riboflavin production consisting of *L. plantarum* strain GCK and *Δrib4:rib5*. (B) Twelve identical populations were evolved in parallel in 96well plates. The growth of each community improved similarly for every replicate, based on Optical Density measurements.

2.2.1.2 Laboratory evolution and determination of the riboflavin secreting phenotype

In order to start the evolution process we combined our chosen *L. plantarum* strain and the auxotrophic *S. cerevisiae* with the same starting optical densities (each with an OD_{600} equal to 0.01) and we subjected 12 identical communities in parallel to a serial transfer experiment (Figure 2.5A). The duration of the experiment was a period of 3 months and in total each of the twelve replicates was transferred 24 times. Since the readout is for the community, and not monocultures, it is difficult to calculate the exact number of generations that passed during the period of the serial transfer experiment. For this reason, we used the difference of measured optical density between the transfers to estimate the number of generations that would have passed if we had one of the two species in monoculture. Based on this estimation the average number of generations that passed for the twelve populations was equal to 160. Next, we determined whether all populations behaved in the same way during the evolution, or if we could see distinct patterns, that would exhibit the dynamic nature of the metabolic cross-feeding interactions.

During the period of evolution, the optical density of the twelve populations increased on average more than sevenfold, with visible oscillations between the transfers. Another aspect that showcases the improvement of fitness that occurred throughout the evolution process is the observed decrease of the time between the transfers. According to our original experimental design the transfers were performed when there was no more observable increase of growth, which was specified for each replicate. Initially the populations were transferred every 5 days, with the time frame decreasing to 4 days after 4 passages and to 3 days after 6 passages. Culture time of 3 days was kept for the rest of the evolution process to ensure that the populations were completely grown, since the secretion of riboflavin is increased during stationary phase (von Canstein *et al.*, 2008). It is worth to mention that by the end of the experiment one population collapsed completely (Figure 2.5A), enhancing the observation that the interactions between the partners of the community remain dynamic during the evolution experiment. After the completion of the serial transfers single colonies of *L. plantarum* were isolated from the 11 remaining final populations for further characterisation. Specifically, the amount of secreted riboflavin was evaluated in cell free culture supernatant by measurement of fluorescence in a plate reader (440,520 nm). In total, we isolated 134 strains and cultured them in CDM34R- supplemented with amino acids (CDM46R-) for the same duration as the period between transfers, which at the end of the ALE experiment equalled to 72 h. Around 60% of those isolates exhibited an improved phenotype, with the $\log_2$fold change varying between 0.15 and 2.5 (Figure 2.6). Apart from strains that exhibited an improved phenotype, a significant proportion of the final populations exhibited 'cheating' behaviour. We use the term cheaters to characterise strains that performed worse than the parental strain and yet they remain as part of the community during the evolution process, without contributing their share of "common goods". Measuring the amount of secreted riboflavin with fluorescence allowed us to efficiently screen a large number of isolates and to evaluate the total percentage of the populations that were improved through the serial transfer experiment. Nevertheless, we decided to test further with analytic methodologies, some of the isolates that exhibited promising phenotype, aiming to identify with increased accuracy the changes that occurred in the secreting phenotype.

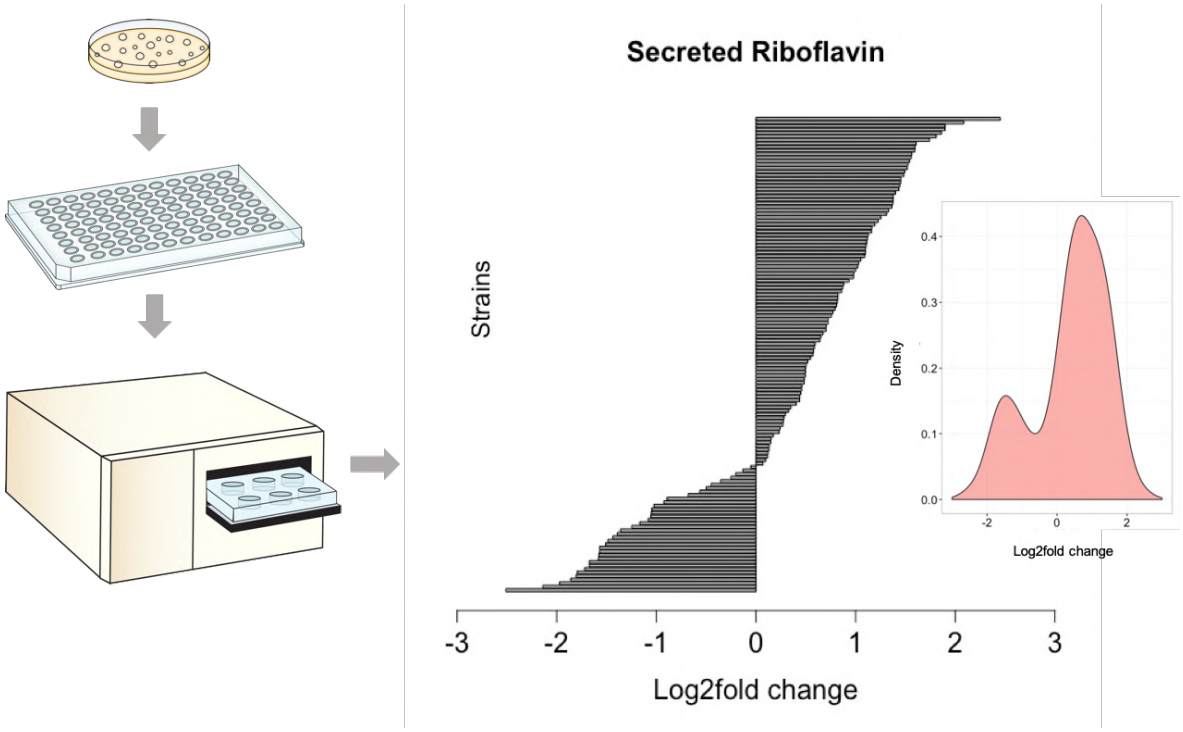

Figure 2.6. Representation of the chosen methodology to screen evolved bacterial clones for their ability to secrete riboflavin. The final populations from the community ALE were platted on selective medium for *L. plantarum* and singles colonies were inoculated in CDM46R-. Cell free supernatant was transferred to a new plate and then the fluorescence intensity was measured (440/520 nm). The majority of the strains exhibited an improved phenotype, based on the $\log_2$fold change of the measured fluorescence, compared to the parental strain.

2.2.1.3 Evaluation of riboflavin biosynthetic phenotype through UPLC

The riboflavin producing ability of the parental and the evolved strains was assessed through ultra performance liquid chromatography (UPLC) in bacteria monocultures at late exponential phase. Firstly, we aimed to determine the levels of riboflavin secreted in the environment. The extracellular fold difference detected between the parental strain and the evolved isolates was on average 20% more for strains B4 and E6, which were the best performing strains (Figure 2.7B). Next, after estimating the difference in secretion levels we tested whether all the available riboflavin is being secreted, by determining the intracellular levels of riboflavin. Indeed, higher riboflavin changes were observed intracellularly in the evolved strains with all evolved strains exhibiting increased amounts of riboflavin (Figure 2.8B). Strain B4 exhibited the highest change in comparison to the parental, which was equal to a 3.5-fold change (Figure 2.8C). Riboflavin is the precursor molecule for the Flavin coenzymes FMN and FAD and thus we decided to take into consideration the amounts of these compounds as well, in order to have a clearer view on the way evolution acted upon the cells during the ALE experiment. While we were not able to detect FMN in any of the

extracellular samples, the two best riboflavin secretors B4 and E6 were secreting more FAD than their parental (1.1-1.2-fold change) (Figure 2.7A and C).

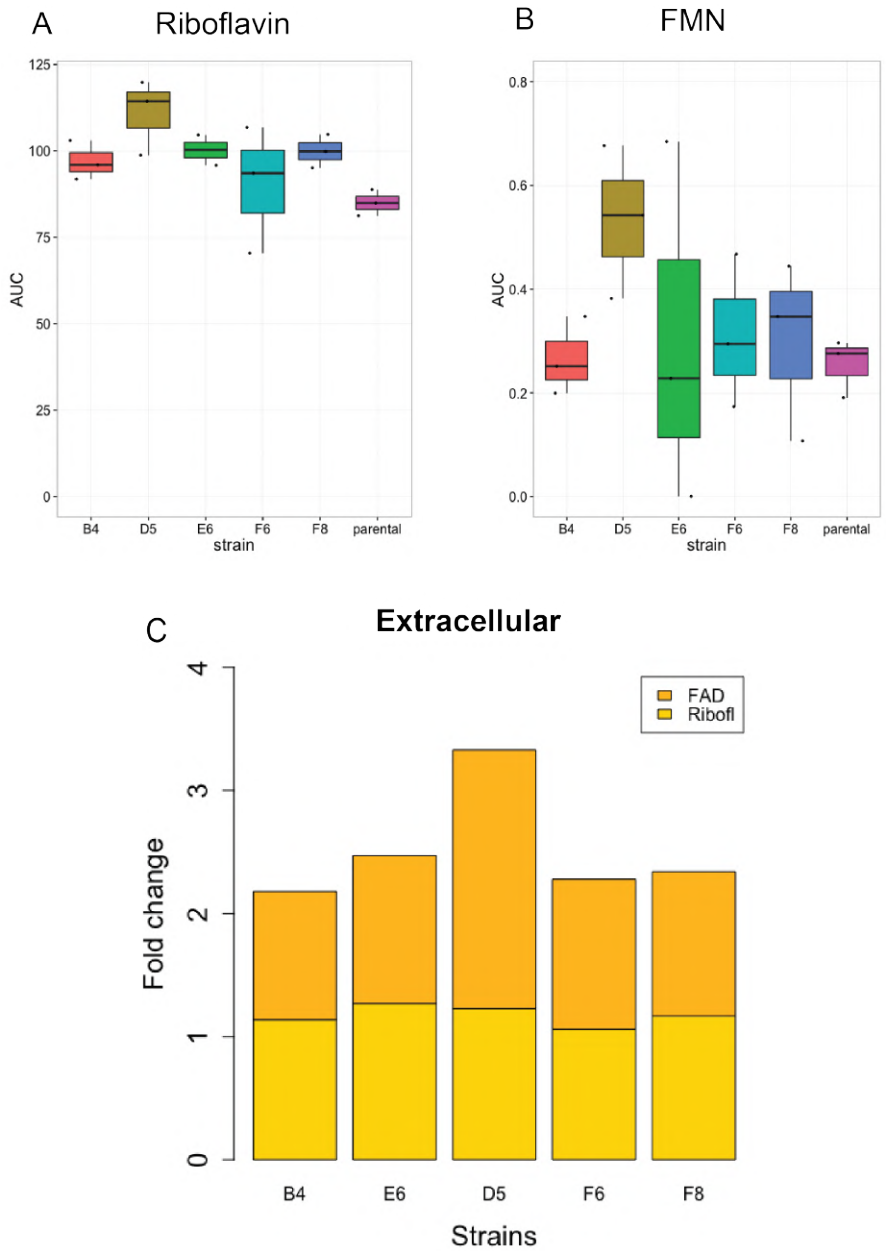

Figure 2.7. UPLC analysis of extracellular samples from the parental *L. plantarum* strain and five evolved strains. (A) In the extracellular samples of all the tested strains, the co-factor FAD which is the active form of riboflavin, was present. In the majority of the tested strains FAD was present in higher amount than in the parental strain, based on the area under the curve (AUC) of the eluted peak. (B) Similar trends were observed for riboflavin, which was also present in free form in the medium, again based on the AUC. In panels A and B each value of the replicates is represented by a black dot. (C) The combined amount of both compounds detected in the samples of the evolved strains was 2-3 times more than in the samples from the parental strain. The bar plot represents fold changes compared to the parental strain based on the average values of the triplicates shown in panels A and B.

In contrast to the extracellular samples, in the intracellular samples we were not able to detect FAD, but the levels of FMN were 2.7 and 3.4 times more for strains E6 and B4

respectively (Figure 2.8A). These results hint that during the evolution process strains arose in the population which could secrete higher amounts of the more readily to use FAD, which perhaps can be taken up by the auxotrophic yeast cells and used up faster in their metabolism. However, there is not solid evidence in literature concerning import of extracellular FAD in *S. cerevisiae*, while in *Pichia guilliermondii* the riboflavin permease was not able to transport either FMN or FAD (Abbas and Sibirny, 2011). Also, it seems that the bacterial strains keep high intracellular pools of riboflavin and FMN, which are either periodically secreted or primarily used for their needs and the surplus is being secreted, having been shown that increased amounts of FMN lead to increased accumulation of riboflavin (Abbas and Sibirny, 2011). Even though we were able to identify interesting differences between the parental and the evolved strain concerning the metabolism of riboflavin, we decided to further validate these differences by liquid chromatography-mass spectrometry (LCMS) and at the same time clearly identify the compounds of interest.

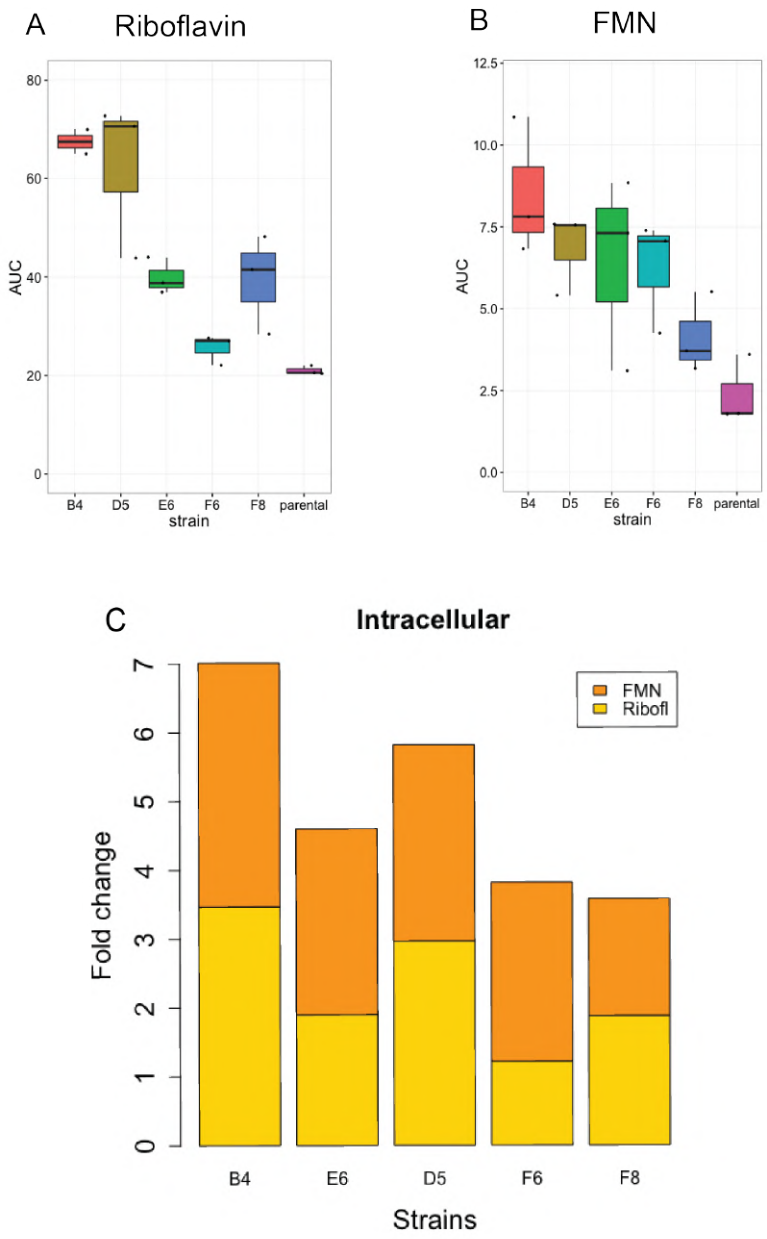

Figure 2.8. UPLC analysis of intracellular samples from the parental *L. plantarum* strain and five evolved strains. (A) The amount of detected riboflavin was increased in the evolved strains, while (B) freely available FMN was detected in the samples of both parental and evolved strains. In panels A and B each value of the replicates is represented by a black dot. (C) The amount of riboflavin was up to 3 times higher in the evolved strains and FMN was up to 3.5 times more. The bar plot represents fold changes compared to the parental strain based on the average values of the triplicates shown in panels A and B.

2.2.1.4 Confirmation of the riboflavin over-producing phenotype through LCMS

When we analysed the bacterial monoculture by LCMS we were able to detect even higher changes in the levels of the extracellularly secreted riboflavin compared to the measurement by UPLC (Figure 2.9A and B). Specifically, the good secretor B4 exhibited a 2.7-fold improvement, while strain E6 was secreting on average 6 times more riboflavin than the parental strain. Similarly to the use of UPLC, we were unable to detect FMN in any of

the samples, and the differences in the levels of FAD are close to the first estimations (fold-change of 1.1 to 1.2) (Figure 2.9C). The observed differences between the two methods may be connected to the increased sensitivity of the LCMS instrument and to the potential degradation of riboflavin during the sample preparation for the UPLC analysis. Interestingly, two more strains that were tested were not releasing higher amounts of riboflavin than the parental, but still the detected amount of FAD was on average 1.6 times higher than the parental (Figure 2.9C).

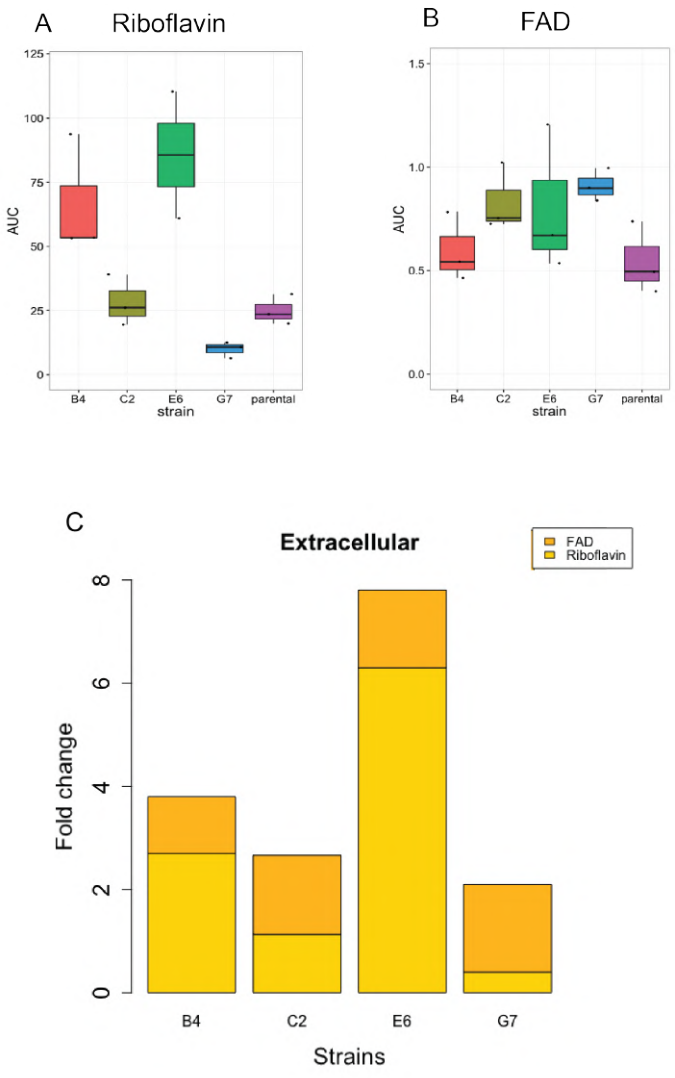

Figure 2.9. LCMS analysis of extracellular samples from the parental *L. plantarum* strain and four evolved strains. (A) In the extracellular samples of all the tested strains FAD was present. In the majority of the tested strains it was present in higher amount than in the parental strain, based on the area under the curve (AUC) of the eluted peak. (B) Similar trends were observed for riboflavin, which was also present in free form in the medium, again based on the AUC. In panels A and B each value of the replicates is represented by a black dot. (C) The combined amount of both compounds varied between strains, with two strains secreting above 2 times more riboflavin than the parental. Both compounds were identified based on their mass transition. The bar plot represents fold changes compared to the parental strain based on the average values of the triplicates shown in panels A and B.

Finally, we analysed intracellular samples by LCMS and we were able to detect more noticeable differences between the evolved strains and their parental, than described for the UPLC analysis. In more detail, strain B4 exhibited a 5.5-fold increase in its riboflavin levels (Figure 2.10C), with the levels of FMN (Figure 2.10A) and FAD (Figure 2.10B) being increased by 10 and 2.7 times respectively. Concerning strain E6, riboflavin levels were increased by 2.8 times, with the levels of FMN and FAD being strongly increased as well (6.2 and 8.3-fold). Interestingly, strain G7 that was releasing less riboflavin to the environment than the parental strain, has stored intracellularly 2 times more riboflavin and more than 4 times more FAD and FMN (Figure 2.10D). The observed differences between the amounts of riboflavin, FMN and FAD may be associated with the state of the cell cycle of the harvested cells, as we chose to perform our experiments with cells in the late exponential phase. Another potential explanation is that riboflavin is being recycled between FAD and FMN, in order to overcome a negative feedback loop due to increased amounts of these compounds, while periodically being secreted to the environment. In conclusion, we were able to detect improved production and secretion of riboflavin in the majority of the tested evolved strain, independently of the method used to estimate the levels of the vitamin and its related compounds. However, we still had to evaluate the effect that the overproducing phenotype had on the general cell fitness.

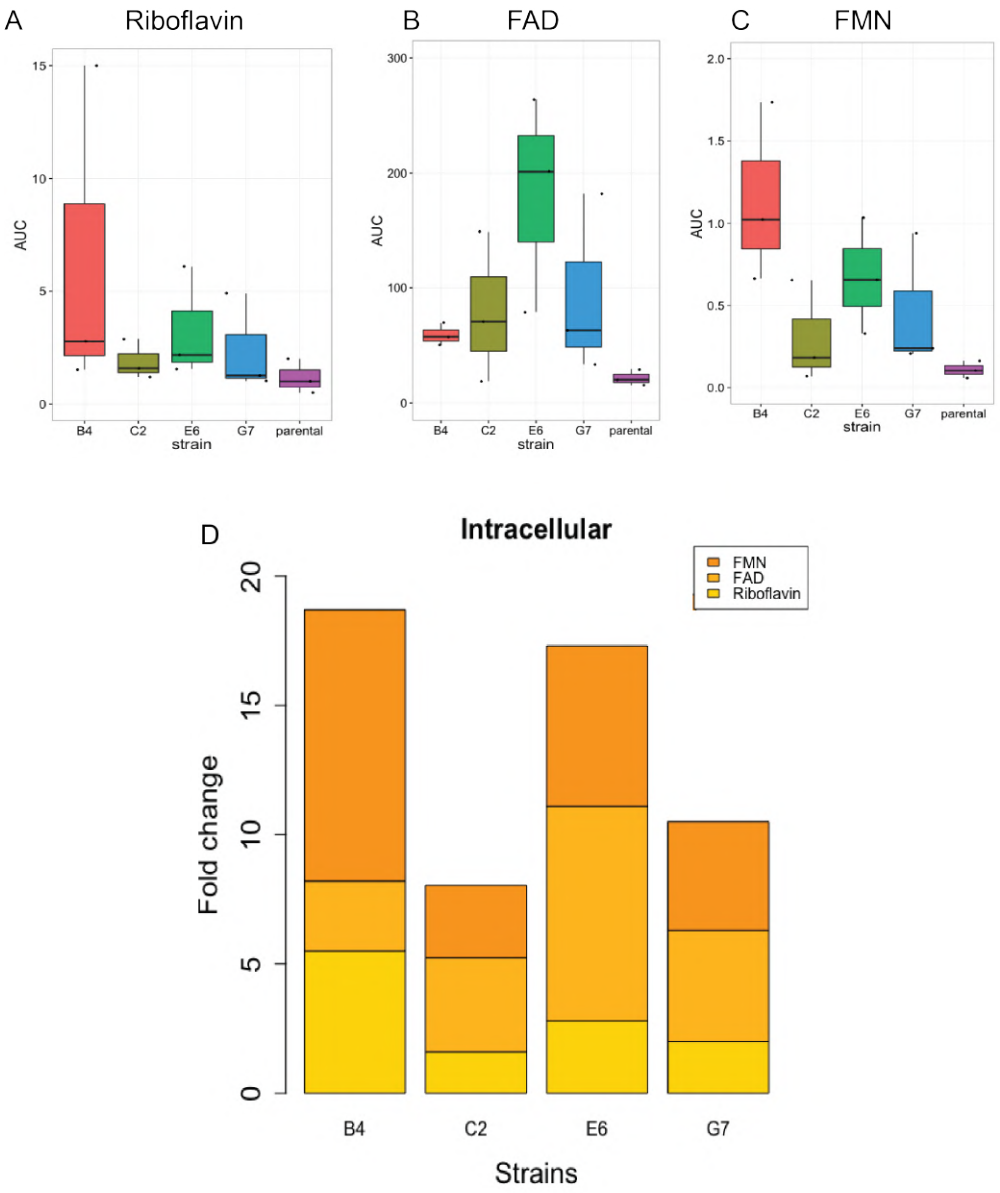

Figure 2.10. LCMS analysis of intracellular samples from the parental *L. plantarum* strain and four evolved strains. All three target compounds were present in both the samples of the parental and the evolved strains. (A) FMN, (B) FAD and (C) riboflavin were detected in increased amounts. In panels A, B and C each value of the replicates is represented by a black dot. (D) Riboflavin was increased up to 5 times in the evolved strains, with total amounts ranging between the evolved strains. The bar plot represents fold changes compared to the parental strain based on the average values of the triplicates shown in panels A, B and C.

2.2.1.5 Summary

All the above described values are summarised for strains B4, C2, E6 and G7 in Table 2.1. This table contains information about extra and intra-cellular riboflavin levels and it can be appreciated that independently of the chosen method (fluorescence measurement, UPLC, or LCMS) the observations are correlated, while the accuracy of detection is increasing with the complexity of the method. As a result, any of these methods can be successfully applied for the identification of strains with improved secreting capabilities, while exhibiting separate strengths and weaknesses. In short, fluorescence measurement can be used easily to screen a

large number of isolates, but it offers limited amount of information. On the other hand, UPLC and LCMS offer higher sensitivity, but sample preparation is more laborious and it is not easy to test many samples, due to increased cost of machine operating time and consumables.

Table 2.1. Summary of the extra and intra-cellular riboflavin levels. The values of four evolved strains based on different detection methods are described. The calculation is based on the foldchange comparison with the parental strain (N/A: not available).

Strain	*Detection Method*	*Sample*	*Fold change*
B4	Fluorescence	Extracellular	3.1
	UPLC	Extracellular	1.14
	LCMS	Extracellular	2.67
	Fluorescence	Intracellular	N/A
	UPLC	Intracellular	3.47
	LCMS	Intracellular	5.52
C2	Fluorescence	Extracellular	1.04
	UPLC	Extracellular	N/A
	LCMS	Extracellular	1.09
	Fluorescence	Intracellular	N/A
	UPLC	Intracellular	N/A
	LCMS	Intracellular	1.61
E6	Fluorescence	Extracellular	1.13
	UPLC	Extracellular	1.27
	LCMS	Extracellular	6.34
	Fluorescence	Intracellular	N/A
	UPLC	Intracellular	1.9
	LCMS	Intracellular	2.79
G7	Fluorescence	Extracellular	3.02
	UPLC	Extracellular	N/A
	LCMS	Extracellular	0.39
	Fluorescence	Intracellular	N/A
	UPLC	Intracellular	N/A
	LCMS	Intracellular	2.04

Detection of the related compounds FMN and FAD requires the use of either UPLC or LCMS, with LCMS being able to detect better the differences between strains. These results are being summarised in Table 2.2.

Table 2.2. Summary of the extra and intra-cellular levels of FMN and FAD. The values of four evolved strains based on UPLC and LCMS detection are described. The calculation is based on the fold change comparison with the parental strain (N/A: not available, N/D: not detected).

Strain	*Compound*	*Detection Method*	*Sample*	*Fold change*
B4	FMN	UPLC	Extracellular	N/D
	FAD	UPLC	Extracellular	1.19
	FMN	LCMS	Extracellular	N/D
	FAD	LCMS	Extracellular	1.09
	FMN	UPLC	Intracellular	3.54
	FAD	UPLC	Intracellular	N/D
	FMN	LCMS	Intracellular	10.5
	FAD	LCMS	Intracellular	2.7
C2	FMN	UPLC	Extracellular	N/A
	FAD	UPLC	Extracellular	N/A
	FMN	LCMS	Extracellular	N/D
	FAD	LCMS	Extracellular	1.52
	FMN	UPLC	Intracellular	N/A
	FAD	UPLC	Intracellular	N/A
	FMN	LCMS	Intracellular	2.78
	FAD	LCMS	Intracellular	3.6
E6	FMN	UPLC	Extracellular	N/D
	FAD	UPLC	Extracellular	1.19
	FMN	LCMS	Extracellular	N/D
	FAD	LCMS	Extracellular	1.46
	FMN	UPLC	Intracellular	2.67
	FAD	UPLC	Intracellular	N/D
	FMN	LCMS	Intracellular	6.21
	FAD	LCMS	Intracellular	8.32
G7	FMN	UPLC	Extracellular	N/A
	FAD	UPLC	Extracellular	N/A
	FMN	LCMS	Extracellular	N/D
	FAD	LCMS	Extracellular	1.66
	FMN	UPLC	Intracellular	N/A
	FAD	UPLC	Intracellular	N/A
	FMN	LCMS	Intracellular	4.26
	FAD	LCMS	Intracellular	4.26

In summary, with our experiments so far, we showed that it is possible to create a stable mutualistic microbial community in specific environmental conditions. Since this community is stable, we could use it for laboratory evolution for a prolonged period of time at the end of which we isolated bacterial clones which exhibited the ability to secrete increased amounts of riboflavin. We confirmed the phenotypic improvement with three different methods, which exhibit different advantages and disadvantages, thus each one of them should be used dependently on the scientific question at hand. Having confirmed that

the secretion capabilities had improved during the evolution process, we moved on to characterise further the evolved isolates.

2.2.2 Fitness comparison of the parental and evolved strains

2.2.2.1 Comparison of bacterial fitness – monocultures in supplemented CDM

As presented, our ALE design links increased production of a desired compound to the improved fitness of the whole community, however the effect on the fitness of individual strains was not yet examined. Firstly, we investigated how the fitness of the evolved strains is different from their parental strain. For this purpose, growth kinetics of the parental and the evolved strains was performed in CDM46R- (Figure 2.11A). Under these conditions all strains grew to similar optical density, however the growth rate varied between different isolates, with some exhibiting a slower growth rate than the parental strain, while others had almost identical growth rate (Figure 2.11B). It is possible that the lower growth rate is associated with the increased fitness cost of producing metabolically expensive compounds, as it is observed when increased amounts of riboflavin are being produced. One such example is strain B4, which is a good secretor (Table 2.1) that reached the same maximal OD_{600} as the parental, but showed a 33% decrease in growth rate. On the other hand, strain C2 that secreted more FAD than riboflavin (Tables 2.1 and 2.2) had about 12% decrease in growth rate compared to the parental strain.

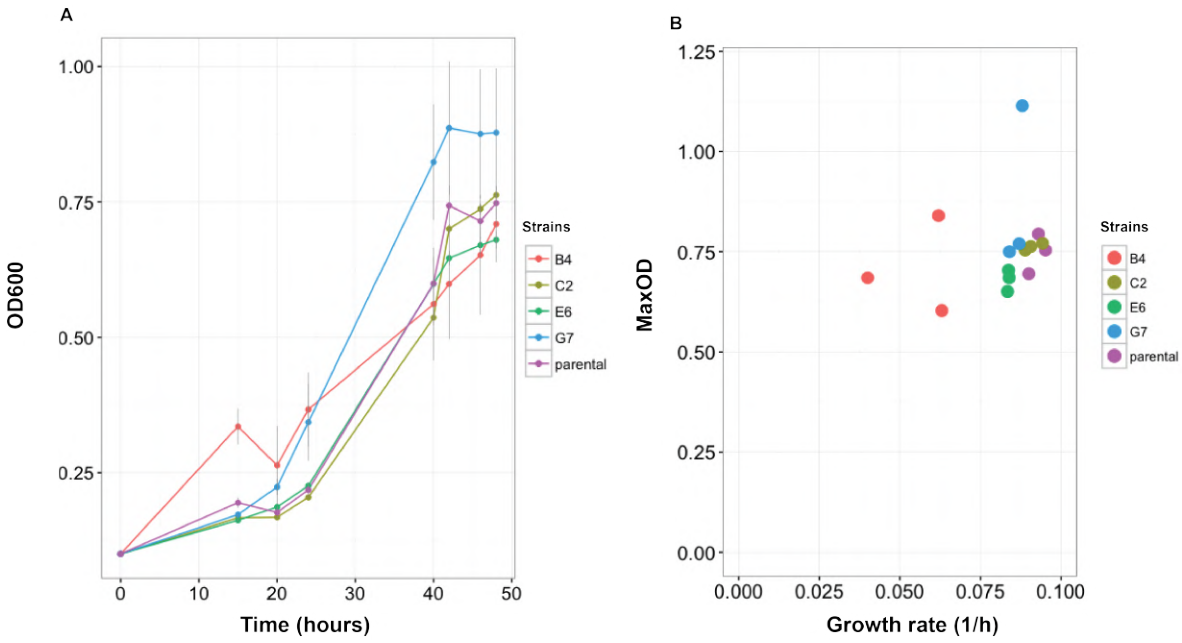

Figure 2.11. Comparison of bacterial monocultures. (A) Growth kinetics of the parental strain and four evolved strains in CDM46R-. In the figure the average value of biological triplicates is represented. (B) Scatter plot of the maximum optical density against the calculated growth rate for each of the replicates from panel A.

2.2.2.2 Estimation of fitness on community level – interspecies interactions

Since the evolved strains as monocultures were not able to grow better than the parental, we set out to investigate whether the general fitness of co-cultures improved during evolution. For this reason, different communities were recreated in CDM34R- and the growth was followed over time. The communities consisted of the parental riboflavin auxotrophic Δ*rib4:rib*5 *S. cerevisiae* and either the parental *L. plantarum* strain, or one of the evolved strains (Figure 2.12A). Communities that included evolved strains were able to reach higher maximal optical density, without changes in growth rates. Moreover, we engineered the parental yeast strain to constitutively express a red fluorescence protein, which we used to sort the two fractions of our population through cell cytometry (FACs). Analysis of different time points along the growth of the co-cultures showed that in the early stages of the growth all bacterial strains support a similar number of yeast cells, but after 72 h, riboflavin producing strains B4 and E6 (Table 2.1) were able to support 1.5 and 4 times more yeast cells respectively (Figure 2.12B).

Interestingly strain C2 which secretes mainly FAD (Table 2.2), supported yeast cells better early during the growth of the co-cultures, specifically at 23 h. At that time point the number of yeast cells was double compared to the parental and the rest of the evolved strains. Still at 72 h it supported 2 times more yeast cells than the parental strain, a number which is comparable to those of strains B4 and E6 (Figure 2.12B). Imaging the same co-cultures by bright field and fluorescence microscopy, showed how the size of the aggregates that form between bacteria and yeast cells was increasing over time. After 23 h of co-culture, we could only detect single yeast cells that were in close proximity with a few bacterial cells. When 72 h had passed, more and more yeast cells were observed to be in close proximity with bacterial aggregates forming between them, while after 84 h had passed, these aggregates had expanded, forming complex structures (Figure 2.12B). These experiments show how the evolution process worked at community level, rather than at the level of single individuals, resulting to bacterial strains which exhibit an over-secreting phenotype arising in the population. These strains produce costly compounds that support their mutualistic partner, even though at first glance increased production negatively affected their fitness under well supplemented conditions.

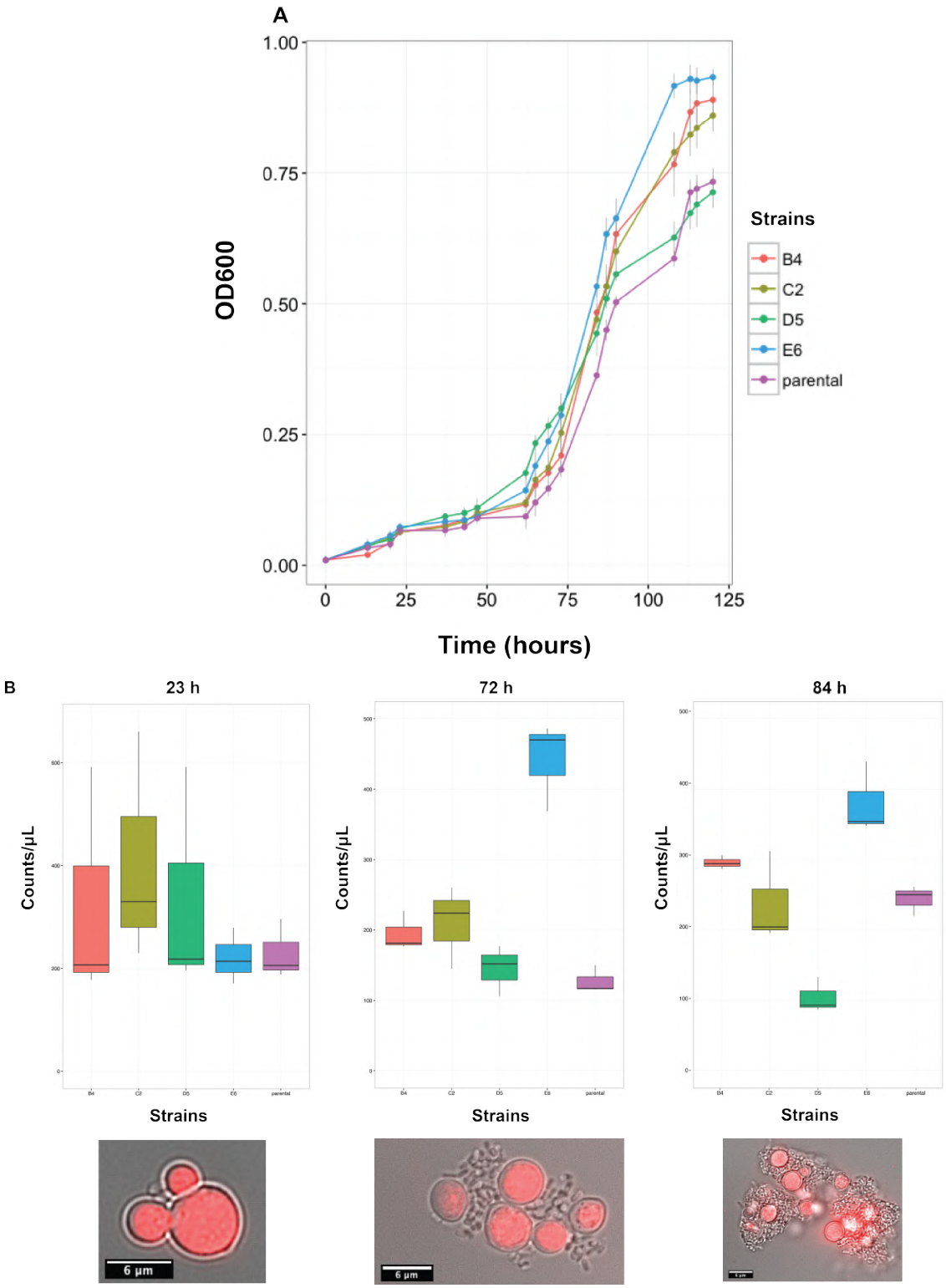

Figure 2.12. Comparison of relative fitness at community level. (A) Growth kinetics of the parental strain and four evolved strains co-cultures with auxotrophic *S. cerevisiae* in CDM35R-. In the figure the average values of biological triplicates are represented. (B) Representation of the number of RFP positive events counted per µL with FACs. Each time point was collected during the kinetics represented in panel A. Below each boxplot is a representative image of a fixed sample from the co-culture of strain E6, as seen in a widefield microscope.

2.2.2.3 The cross-feeding interactions are retained with conditioned media

The next step was to evaluate whether conditioned medium from the parental strain and the evolved isolates can support the growth of auxotrophic *S. cerevisiae* cells in the same or in a different way. So far, we had seen that the desired compound was secreted in the environment even without the presence of yeast cells and that the yeast population grew when it was co-cultured with evolved strains, forming strong aggregates. We set out to test if the improved yeast phenotype persists in the absence of bacterial cells and these aggregates. For this purpose, auxotrophic yeast cells were inoculated in conditioned media from the chosen bacterial strains, without any further supplementation. Cultures of parental auxotrophic *S. cerevisiae* in conditioned media of evolved isolates were able to reach an optical density 1.5 times higher and a growth rate 1.4 times higher than yeast cells grown in conditioned medium of the parental *L. plantarum* strain (Figure 2.13A). This evidence supports the previous observations that the bacterial cells evolved to better support the growth of yeast cells, while it seems that direct contact between the cells is not necessary as the yeast can uptake the compounds directly from the environment.

Moreover, the ability of the parental and the evolved LAB strains to use conditioned medium from wild type *S. cerevisiae* was evaluated. In a similar fashion, bacterial cells were inoculated in the conditioned medium of the WT S90 yeast strain. Evolved strains were able to reach a higher maximal OD (Figure 2.13B), a fact that suggests that during the process of evolution they became more efficient in amino acid utilisation and thus balanced the fitness deficit of riboflavin overproduction. For example, strains B4 and E6 reached 1.5 times the maximum OD_{600} values, but the observed growth was quite slow with all the strains reaching their maximum OD_{600} after 100 hours of culture. In comparison, when grown in CDM46R- the required time for reaching the maximum OD_{600} is around 90 hours and the absolute value is 10 times higher.

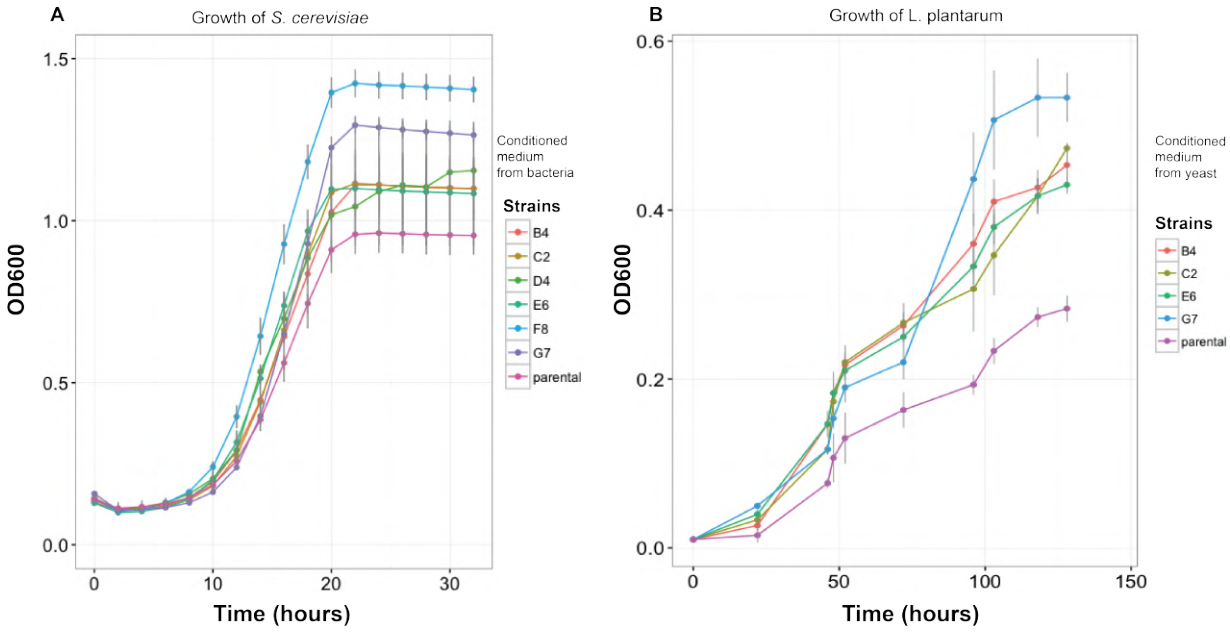

Figure 2.13. Comparison of relative fitness in conditioned media. (A) Growth kinetics of the parental *Δrib4:rib5 S. cerevisiae* strain in conditioned medium of the parental and evolved bacteria strains. In the figure the average value of biological replicates is represented. (B) Growth kinetics of the parental and evolved bacteria strains in conditioned medium from wild-type *S. cerevisiae*. In the figure the average values of biological triplicates are represented.

2.2.2.4 Estimation of cellular fitness under limited nitrogen conditions

When all the strains are cultured in fully supplemented CDM, no significant differences in growth are noticeable, but the evolved strains grew better in yeast conditioned medium, hinting towards better growth under nitrogen limited conditions. To test this hypothesis, we cultured the selected strains in CDM46R-, but the concentration of the supplemented amino acids was diluted 8 times to resemble amino acid limiting conditions (Figure 2.14A). Under these conditions the evolved isolates were able to reach a higher maximal OD_{600}, also the growth rate of the parental strain decreased while all the evolved strains remained stable or their growth rate increased (Figure 2.14B). More specifically, the growth rate of the parental *L. plantarum* decreased by 17% and the maximum OD_{600} by 22%, when it was grown in the presence of reduced amino acid concentration. In contrast, the riboflavin over secreting strain B4 (Table 2.1) increased its growth rate by 40% and its maximal OD_{600} by 8%, while FAD secreting strain G7 (Table 2.2) retained the same growth rate with a subsequent decrease of 3% on its maximal OD_{600}.

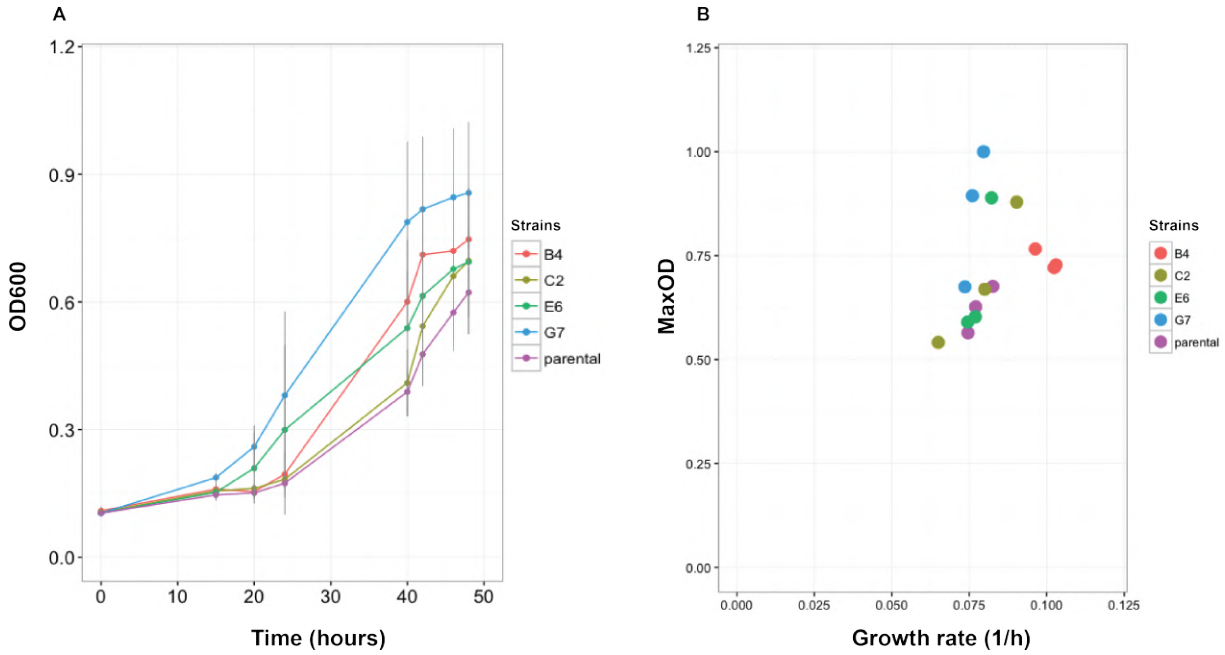

Figure 2.14. Comparison of bacterial monocultures under limited nitrogen conditions. (A) Growth kinetics of the parental strain and four evolved strains in CDM46R- with decreased amount of supplemented amino acids. In the figure the average value of biological triplicates is represented. (B) Scatter plot of the maximum optical density against the calculated growth rate for each of the replicates from panel A.

2.2.2.5. Summary

With this set of experiments, we were able to show that evolved strains that exhibit an improved riboflavin secreting phenotype were not exhibiting improved fitness compared to the parental strain under fully supplemented conditions. However, the same strains were able to perform better both in co-culture with auxotrophic yeast and under amino acid limited conditions (diluted amino acids and yeast conditioned medium). As shown in Figure 2.15, increased production of riboflavin seems to be associated with a lower growth rate. However, this correlation is not very strong, especially since the three evolved strains C2, E6 and G7 behaved similarly, even though they differ in their secreting phenotype (Table 2.3). Therefore, it seems that optical density and growth rate can give hints about the effect of riboflavin overproduction on the physiology of the cell, but it is still necessary to look deeper into the regulation of this aspect within the cell metabolism.

Table 2.3. Summary of the riboflavin biosynthetic capabilities and the cell fitness. The amount of produced riboflavin for the parental and four evolved strains, is calculated as the fold change in comparison to the parental strain and maximal growth and growth rate are used as measures of relative fitness.

Strain	*Extracellular Riboflavin (fold change)*	*Intracellular Riboflavin (fold change)*	*Maximal OD_{600} Complete medium*	*Growth rate (h^{-1}) Complete medium*	*Maximal OD_{600} Diluted A.A*	*Growth rate (h^{-1}) Diluted A.A*
parental	1	1	0.74	0.092	0.62	0.075
B4	2.67	5.52	0.68	0.061	0.74	0.1
C2	1.09	1.61	0.76	0.084	0.69	0.07
E6	2.79	2.79	0.68	0.083	0.69	0.078
G7	0.39	2.04	0.878	0.077	0.85	0.076

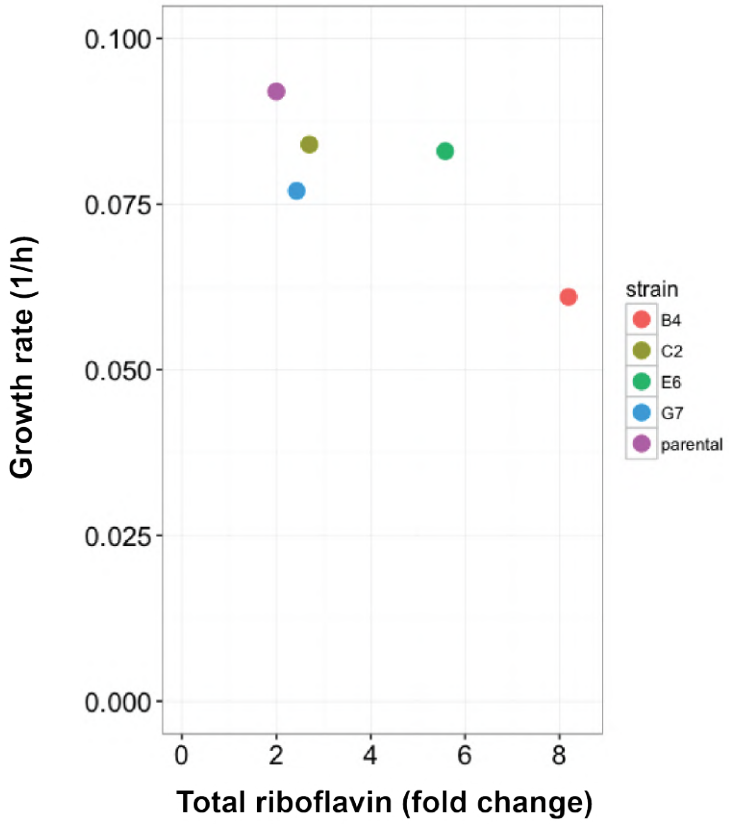

Figure 2.15. Comparison of the growth rate and total amount of detected riboflavin (intracellular and extracellular). Scatter plot of the maximum calculated against the total amount of produced riboflavin calculated as fold change compared to the parental strain.

2.2.3 Genetic and regulatory changes underlying increased vitamin production during coevolution

2.2.3.1 Whole genome sequencing reveals changes in genes connected to transcriptional regulation and transport of amino acids

The next step in our study was to try to elucidate the changes that occurred during the evolution process, which led to the observed improvement of the phenotype. For this purpose, 14 evolved strains and 2 of the final populations from the ALE co-culture experiment were sequenced and compared to the sequence of the parental strain, with the results being summarised in Table 2.4. It is worth mentioning that none of the identified mutations were located in the riboflavin operon, or its upstream regulating area, but we identified single nucleotide polymorphisms (SNPs) along the whole chromosome (Figure 2.16). In the sequenced populations (Table 2.4), we were able to identify mutations in genes that regulate gene transcription, such as a LysR family transcriptional regulator, the transcription aniterminator BgIG and the bifunctional pyr operon transcriptional regulator. Interestingly, SNPs were detected in a number of different genes associated with the uptake of amino acids. For example, in one of the populations, one SNP was detected in 11.7% of the reads, in the gene of one of the carrier proteins of the branched-chain amino acids transport system. Moreover, we identified different SNPs in the genomic sequence of at least 2 different amino acid permeases, with one penetrating the population at 11.2%. Mutations were also detected in a variety of other transporters including transporters of peptides (15%) and transporters of sugars, such as a phospotransferase (PTS, 47%). Another group of related mutations was the group of mutations that affect the morphology of the cell, such as different cell surface proteins, cell wall anchor proteins and the cell shape-determining protein MreB. Finally, we detected SNPs in metabolic enzymes that can be potentially associated with the general metabolism of riboflavin. Such an example is the detection of two separate synonymous mutations that were present in the gene which encodes for the enzyme ribito-5-phosphate dehydrogenase, which catalyses the formation of ribulose-5-phosphate. Along with GTP, ribulose 5-phospate is the second most important precursor for the biosynthesis of riboflavin. Lastly, mutations were identified on enzymes related to glycolysis, such as pyruvate transaminase, a NAD depended malic enzyme, a SNP in the gene coding for a fumarate reductase in strain C2 and the 6-phospogluconate dehydrogenase which is involved in the production of ribulose 5-phospate.

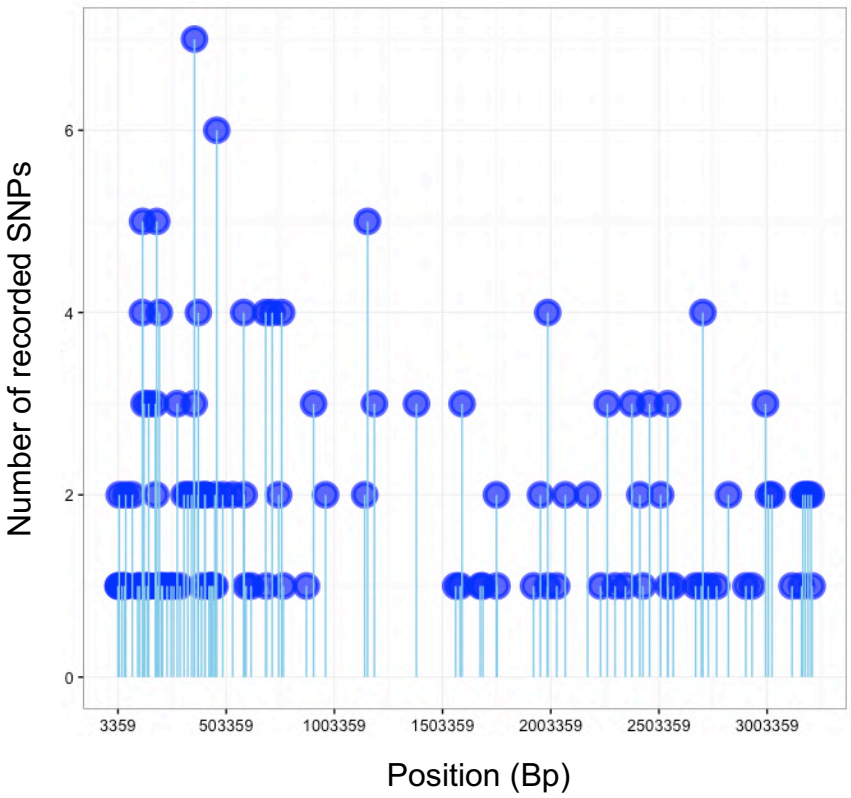

Figure 2.16. Representation of the SNPs distribution along the bacterial chromosome. In the mixed population samples SNPs were identified all along the length of the chromosome (*x* axis), with different genes exhibiting higher or lower number of uniquely identified SNPs (*y* axis).

Table 2.4. List of the identified mutations in the two mixed populations. The mutations listed here can be partially correlated with the chosen ALE conditions and with the described phenotypic differences.

Position (bp)	***Population A (%)***	***Population B (%)***	***Mutation***	***Gene Description***
39059	48.4	-	A192E(GCG→GAG)	PAS domain-containing sensor histidine kinase
178031	20.7	-	T→G intergenic (+51/+149)	LysR family transcriptional regulator
178023	-	16.7	A→T intergenic (+43/+157)	LysR family transcriptional regulator
182721	10.3	-	A→C pseudogene (2641/2847 nt)	Cell surface protein
182516	-	37.4	G→A pseudogene (2436/2847 nt)	Cell surface protein
185126	15	-	T262P (ACC→CCC)	Peptide ABC transporter substrate binding prot
441674	10.3	-	V114G (GTC→GGC)	Serine-tRNA ligase
405205	-	11.1	G4V (GGA→GTA)	Transcriptional regulator

531629	47	-	Q266K (CAA→AAA)	PTS mannose family transporter
533398	-	25.5	R271Q (CGA→CAA)	Transcription antiterminator BgIG
573497	-	11.9	G→T intergenic (-97/+42)	Ribonucleotide-diphospate reductase sub alpha
584288	-	18.7	C→A intergenic (+42/+92)	UDP glucose epimerase/pyruvate transaminase
586563	10.7	-	S206A (TCA→GCA)	ABC transporter permease
725060	18.6	-	Q271H (CAA→CAC)	RNA degradosome polyphospate kinase
739003	22.8	-	S314C (TCT→TGT)	Biotin-(acetyl-CoA-carboxylase) ligase
745865	18	-	A232A (GCT→GCG)	Amino acid permease
967174	9.7	-	C72G (TGT→GGT)	Amino acid permease
967197	-	11.2	Y79* (TAC→TAA)	Amino acid permease
1161131	-	27.4	C→A pseudogene (3738/8007 nt)	Cell surface determining protein
1161185	23.1	-	T→C pseudogene (3792/8007 nt)	Cell surface determining protein
1567624	12.6	-	T→G intergenic (+120/-39)	Bifunctional pyr operon transcriptional regul
1588282	9.4	-	V109G (GTC→GGC)	Cyclic nucleotide binding protein
1597405	8.9	-	L175L (TTG→TTA)	Ribitol-5-phosphate dehydrogenase
1597559	-	6.6	L227L (TTA→GTA)	Ribitol-5-phosphate dehydrogenase
1759183	26.3	-	K351N (AAA→AAC)	GTP pyrophosphokinase
1927917	13.3	-	N68T (AAC→ACC)	GTP binding protein TypA
2464257	-	41.3	C→A intergenic (+138/+140)	LysR family transcriptional regulator/tRNA-Thr
2772422	-	9.9	H33N (CAC→AAC)	Malate dehydrogenase
3030091	11.7	-	V404G (GTC→GGC)	Branched-chain amino acid transport system

Each of the single strains had on average 11 unique mutations. A gene ontology (GO) terms analysis (Figure 2.17A), showed that the most common categories of genes for which we identified mutations included transporters, transcription related proteins, and cell surface

or hypothetical proteins. These results agree with the observations that we made in the sequenced populations. For example, in strain B4 we identified mostly SNPs related with transcription regulation, such as a transcription antiterminator BgID and a TetR family transcriptional regulator. The family of TetR regulators is associated with the regulation of genes that encode for small-molecule exporters (Cuthbertson and Nodwell, 2013), but we cannot predict whether the SNP we identify is resulting in a gain or loss of function. It is worth to mention that strain E6 that exhibited the best secreting phenotype (Table 2.1), had the least number of mutations.

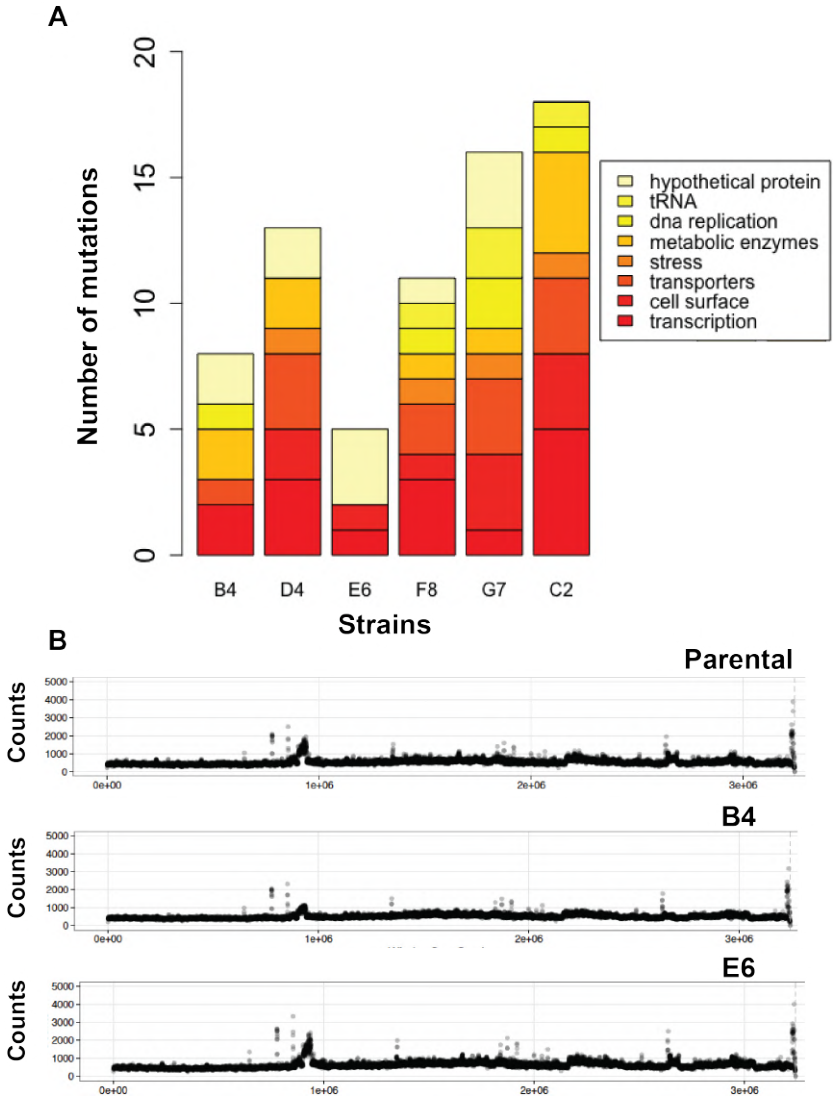

Figure 2.17. Most common mutations detected in single strains. (A) Gene ontology analysis of the unique mutations, detected in the evolved strains. (B) Coverage of the NGS reads along the chromosome, as a method to detect potential insertions, deletion or duplications that occurred during the evolution process.

Apart from the mutations we could detect with use of the algorithm Breseq which can identify SNPs, small INDELs (insertions and deletions) and the emergence of new junctions, we investigated whether any major duplications or deletions had occurred. For this reason,

we tested the coverage of the reads from strains B4 and E6 and compared it to the reads' coverage from the parental strain (Figure 2.17B). The coverage profiles of both evolved strains looked very similar comparing to the profile of their parental strain and no major differences were detected, including in the riboflavin operon. Strain B4 had a slight decrease in coverage in an area between $9x10^5$ and 10^6 bp, an area where numerous metabolic enzymes are located, such as malate dehydrogenase, citrate lyase and L-lactate dehydrogenase. The number of reads in that area is about half for strain B4 compared to strain E6 and the parental, a fact which hints potential adjustment of the general metabolism by some evolved strains, in order to cope with the increased cost of riboflavin production.

2.2.3.2 Proteome analysis indicates increased enzymatic activity towards riboflavin biosynthesis

Through quantitative proteomics on the total proteome of bacterial monocultures that were grown in CDM46R we examined the differences in protein abundance between the evolved strains and the parental strain. An initial principal component analysis (Figure 2.18), showed that the majority of the evolved strains, apart from strain F8, cluster close to the parental and so we can estimate that the general state of the cells did not change excessively through the evolution process, as it was observed in the phenotyping experiments in fully supplemented CDM46R-.

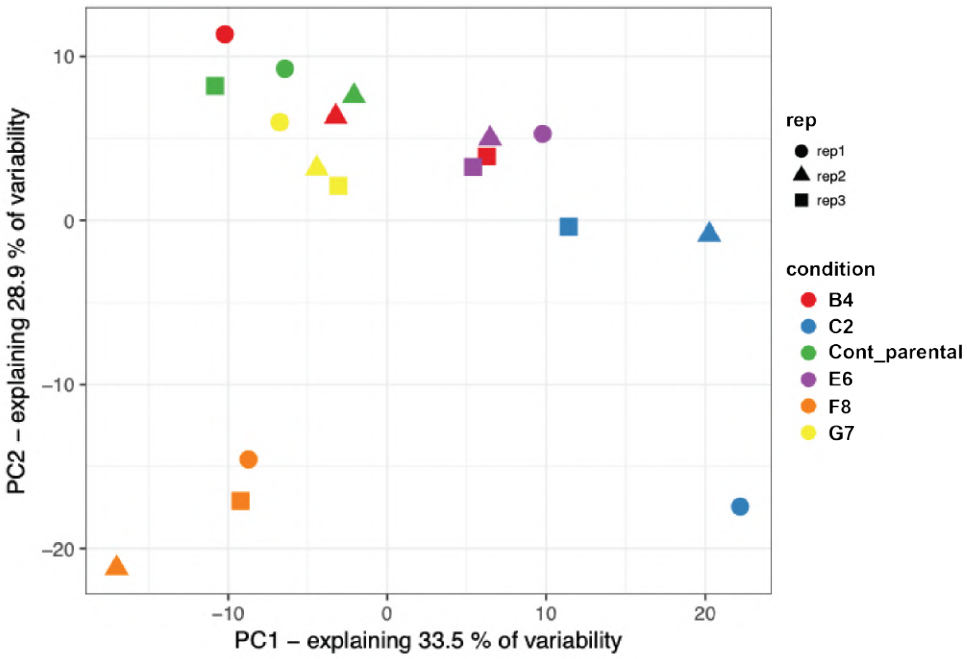

Figure 2.18. Comparison of the bacterial proteome. Principal component analysis of the whole proteome of the parental and the evolved strains. Apart from strain F8, all the other strains are close to their parental strain (Cont_parental).

To identify the most important differences that occurred, we selected those proteins that exhibited at least a log_2fold change of 1 or -1 (p-value < 0.05). Through this process, we ended up with a cluster of 55 proteins which were shared between the majority of the tested strains (Figure 2.19A). This subset of proteins is associated with different aspects of the cell physiology, as it was shown through a Gene Ontology (GO term) analysis (Figure 2.19B). According to the GO term analysis, the proteins of the list most significantly represent the purine biosynthesis pathway (13 keyword counts), like the *de novo* biosynthetic process of IMP (12 keyword counts), as well as the metabolism of purines (17 keyword counts). Other upregulated categories included "nucleotide-binding" (18 keyword counts), ATP-binding (14 keyword counts) and GTP-binding (4 keyword counts), while biosynthesis of riboflavin was represented twice.

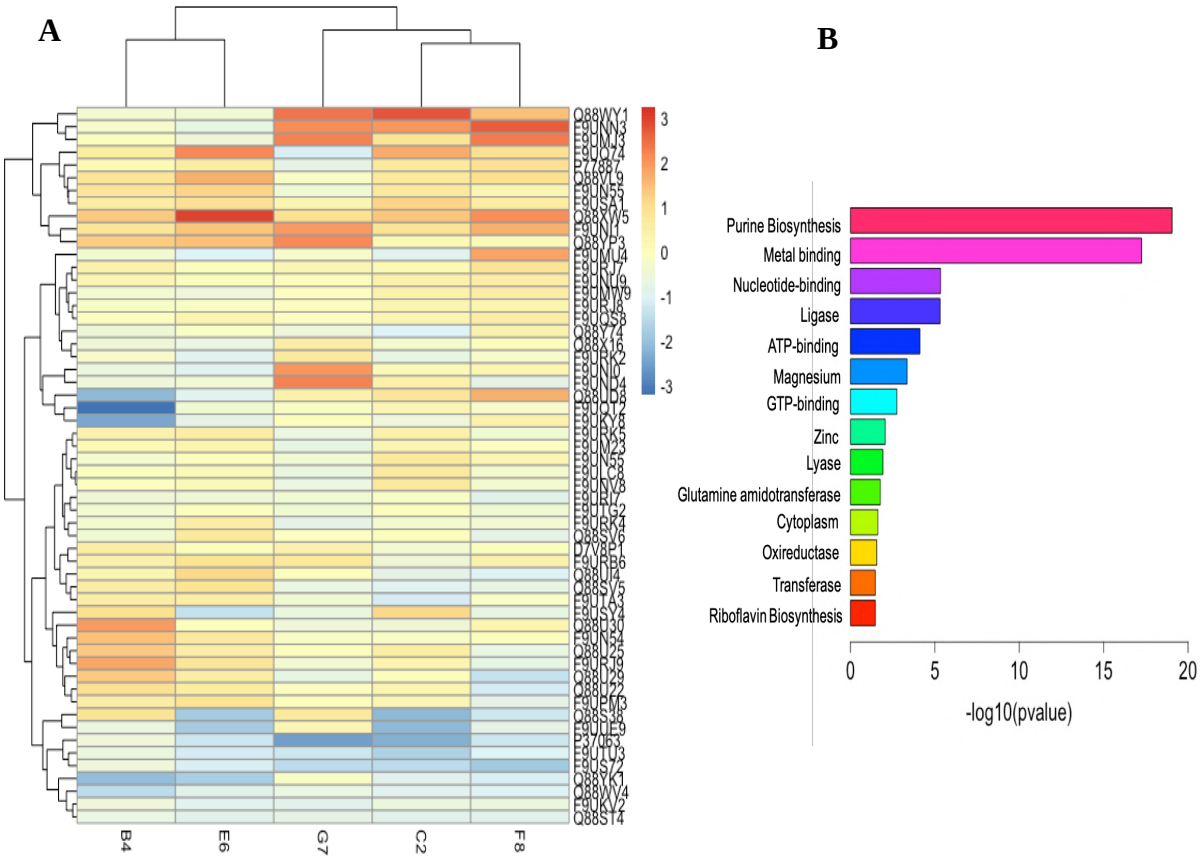

Figure 2.19. Cluster of major differentially regulated proteins. (A) Heatmap of the log_2fold change between evolved strains and their parental, using a cluster of 55 proteins, colour gradient represents the log_2fold change value. The evolved strains exhibit similar expression patterns. (B) Gene ontology analysis of protein set of panel A, using the tool gProfiler (Huang et *al.*, 2009).

In more detail, analysis of the proteomics results showed that all the evolved strains had different protein abundances, related to the biosynthesis of vitamins (Figure 2.20A). First of all, the evolved strains have a higher abundance of the protein RibA (GTP cyclohydrolase-

2), which catalyses the first reaction of the riboflavin biosynthetic pathway. The observed log_2fold change in comparison to the parental strain was ranging from 1.1 in strain G7 to 1.8 for strain E6, which was the best riboflavin secretor (Table 2.1). It seems that increased amounts of this initial enzyme are adequate for increased riboflavin production, since the rest of the enzymes participating in the pathway were increased by smaller amounts. For example, strains G7 and C2 exhibited increased amounts of the protein RibB, with a log_2fold increase of 1.1 and 0.7 respectively, while we detected increased amounts of RibD (0.9 and 0.7 log_2fold increase) in strains B4 and E6. In the vitamin metabolism GTP is also the initial precursor for the biosynthesis of folate, being hydrolysed by the protein FolE (GTP cyclohydrolase-1). For all the tested evolved strains, we observed a decrease in the amount of FolE, which ranged between 0.1 and 0.4 log_2fold change. That hints towards the fact that the selection pressure that we applied during the evolution process was selective for the production of riboflavin and not vitamins in general.

In accordance with the sequencing results that were detected with genome analysis, we were able to identify changes in the level of many transcription regulators, as well as of different transporters. Most of the transcription regulators belong to the same family, with one belonging to the MarR family which is common between for 3 evolved strains. Nevertheless, the variability between strains was quite high, as none of these regulators was upregulated in strain G7, while for strain C2 we found 13 upregulated regulators, belonging to various families. This increase of regulatory proteins hints that during the evolution process, the bacteria cells had to improve the regulation of the majority of the cell functions to efficiently use limited recourses, such as the amino acids secreted by their yeast partner and to cope better with the cost of increased riboflavin production. Concerning transporters, in strain C2 an oligo peptide transporter was increased by 1.4 log_2fold change in comparison to the parental strain, while another one was decreased by 1.8 log_2fold change, a fact that potentially hints towards preferential uptake of amino acids. The hypothesis of amino acid preference is also supported by the observed decrease of a branched-chain amino acid aminotransferase (-2.8 log_2fold change) in strain B4.

The biosynthesis of riboflavin is closely related to the metabolism of nucleotides, more specifically the metabolism of purines. We were able to detect increased levels of purine-related proteins in the proteomes of the evolved strains (Figure 2.20B). First of all, for strain B4 we were able to detect an increase of 1.8 log_2fold change of the protein PurH, a bifunctional purine biosynthesis protein that participates in the *de novo* biosynthesis of inosine monophosphate (IMP). Similarly strain E6 overexpressed a different protein from the

same pathway and additionally the protein PurE, a ribonucleotide mutase. Moreover, strain E6 exhibited increased levels of the protein GuaB, which catalyses the deamination of GMP to IMP and controls the balance between adenine and guanine. In addition, strains C2 and G7 had increased availability of a protein which participates in the purine salvage pathway. It is worth mentioning that in all the evolved strains we recorded increased levels of proteins which are associated with the biosynthesis of pyrimidines were found. For example, the protein dihydroorotate dehydrogenase A (PyrD) was increased by a log_2fold change of 1.4 in strain C2 and 1.2 in strain E6. Even though, the FolE was downregulated, we identified upregulated proteins that are associated with the metabolism of folate and are connected to the metabolism of nucleotides as well. For example, strains C2 and G7 exhibited increased levels of a protein which participates in the formation of pyrimidines from methylated forms of folate (5,10-methylene tetrahydrofolate). A similar protein was upregulated in strain F8, the thymidylate synthase (ThyA), with a 1.2 log_2fold change.

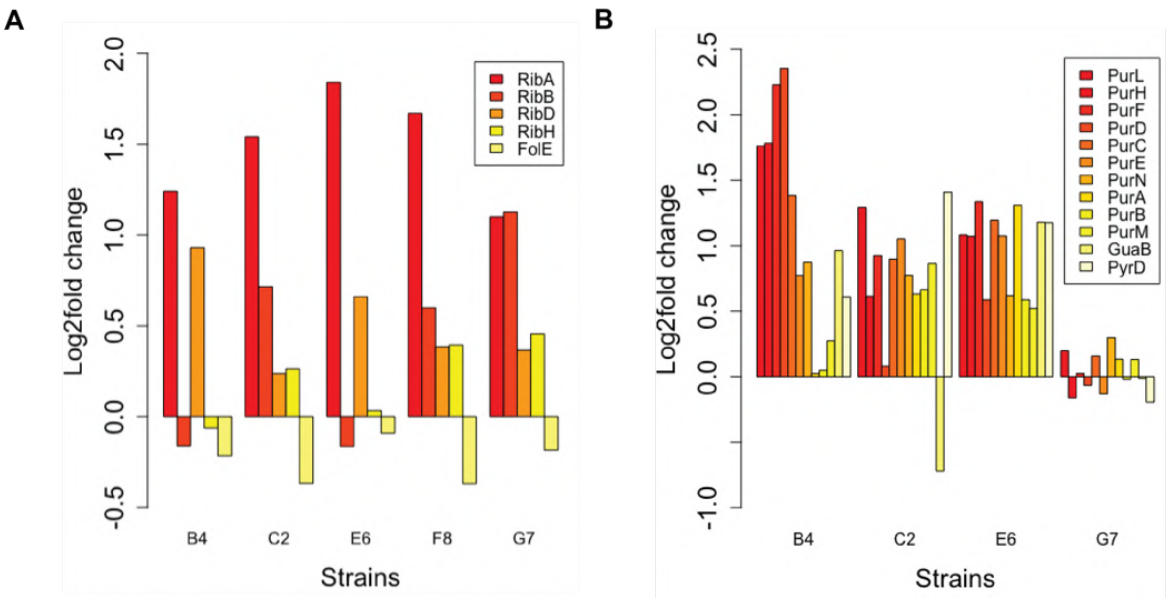

Figure 2.20. Major differentially regulated proteins, correlated with vitamin biosynthesis. Differences in the expression (log_2fold) of (A) proteins involved to the biosynthetic pathway of riboflavin and/or folate, or (B) proteins involved in the biosynthetic pathways of nucleotides.

2.2.3.3. Summary

Mapping of the differently regulated proteins on the metabolic map with iPath3 (Darzi *et al.*, 2018) showed that the evolved strains that were tested, converged towards the regulation of specific pathways, like the riboflavin and the nucleotide pathways, but were different concerning the metabolism of lipids, amino acids and sugars. As seen in figure 2.21 riboflavin overproducing strain E6 has upregulated the pathways that relate to nucleotide metabolism, vitamin metabolism, amino acid metabolism and glycolysis. On the other hand,

the parts of the metabolism that seem to be downregulated include mainly the lipid metabolism, biosynthesis of amino acids and parts of the TCA cycle.

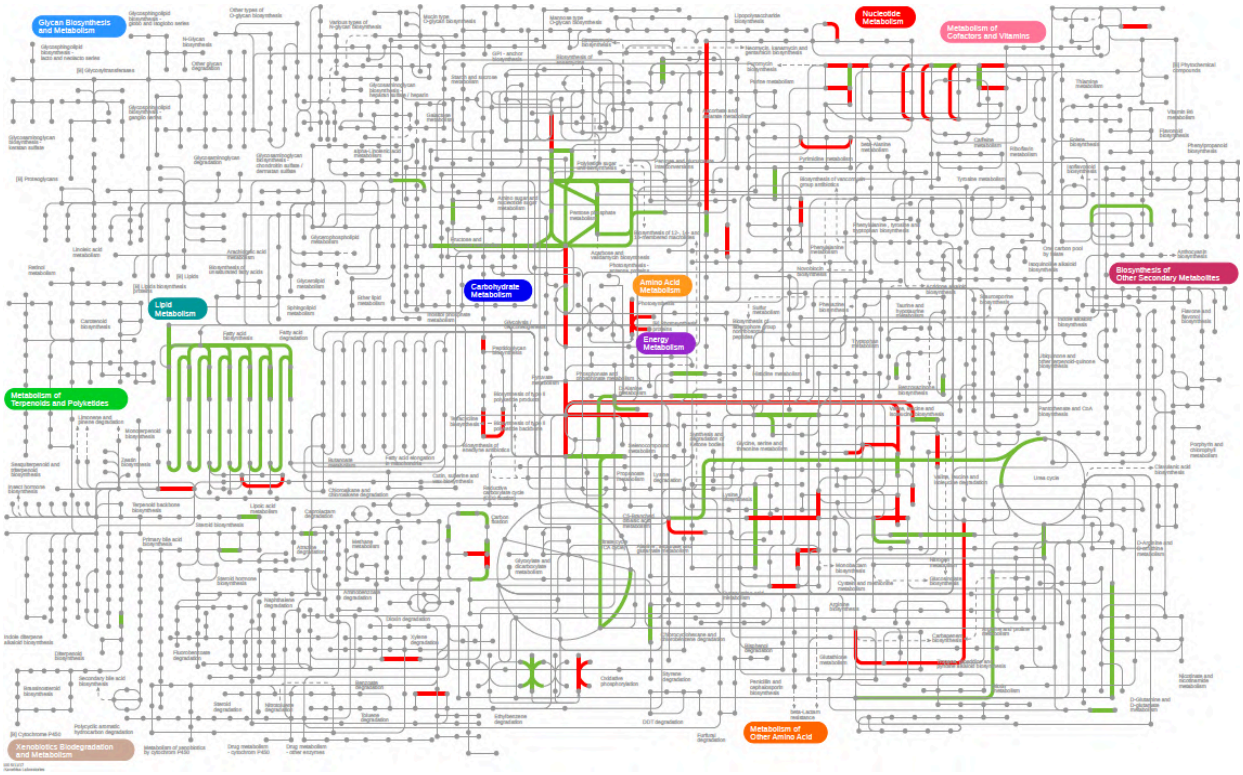

Figure 2.21. Mapping on the metabolism of the differentially regulated proteins. Upregulated (red) and downregulated (green) proteins of the evolved strain E6, correlate to various metabolic activities as seen on the metabolic map. The upper right part represents the pathways of vitamin and nucleotide biosynthesis, in the center is the main carbon metabolism and on the left the lipid metabolism.

Use of NGS and proteomics analysis, gave us a first idea of the changes that occurred in the bacterial cells during the co-culture evolution. First of all, evidence suggests that the bacteria adapted for improved growth in CDM35, by improving their ability to uptake amino acids through increased, or altered activity of transporters. Moreover, we couldn't identify single mutations that would explain the improved capability to produce and secrete riboflavin. However, we identified in the samples of evolved strains, upregulation of enzymes that participate either in the biosynthetic pathway of riboflavin, or pathway that lead to the production of required precursors that are used for the production of riboflavin. Finally, evolved isolates that exhibit similar phenotypes had variable patterns of SNPs and of differentially regulated proteins, meaning that alternative changes can possibly lead to the same outcome. Even though, these experiments allowed us to draw important conclusions, opened questions still remained.

2.2.4 Cell surface traits and spatial structure

One of the major open questions about mutualistic cross-feeding is how it initially emerges in nature and how it establishes over time, since actively investing into the wellbeing of a partner would require an equal gain. At the same time the individuals that promote the growth of their partner need to be protected from the ones that selfishly take advantage of the benefits without investing any resources. It has been suggested (Marchal *et al.*, 2017, Stump *et al.*, 2018) that spatial structure promotes the active metabolite excretion and at the same time limits the benefits, mainly in the form of resources, to the cooperating phenotypes. In our study so far, several SNPs were identified in genes coding for cell surface proteins some of which were associated with the aggregation phenotype. Light microscopy did not allow to calculate the extent of the formed aggregates between evolved and parental strains. For this reason, we set out to evaluate different aspects of the cell surface traits, as a general method to evaluate the physiochemical characteristics of the bacterial cells and their adhesion capabilities (Deepika and Charalampopoulos, 2010). The characteristics of the cell surface of lactobacilli have been used to evaluate the technological importance and the potential probiotic activity of single strains, most commonly natural isolates from food sources, like dairy products (Kotzamanidis *et al.*, 2010).

2.2.4.1 Evaluation of the auto-aggregation and co-aggregation phenotypes

The first characteristic we investigated was the ability of the parental and the evolved strains to form aggregates, both with themselves, i.e. auto-aggregation as well as with yeast cells i.e. co-aggregation. For these tests, we chose to test multiple strains, apart from the ones that we had phenotyped so far, including strains that had performed worse than the parental strain in the initial fluorescence screening (strain A4). However, the observed variability in the formation of aggregates between the strains cannot be directly associated with the levels of riboflavin secretion. In general, after a of short amount of time (1 to 5 h) the evolved strains were aggregating on average 1.5% more than the parental strain. We observed bigger differences after 24 h had passed, as two strains that had not exhibited increased secretion of riboflavin aggregated 3% more than the parental. At the same time point (24 h) the riboflavin overproducing strain B4 aggregated 4% less than the parental. Nevertheless, we cannot hypothesise that aggregation and riboflavin secretion are strongly correlated, due to the fact that strain C2 that exhibited an improved phenotype was able to aggregate 3.5% more than the parental after 24 h (Figure 2.22A).

Similarly, the strains that could auto-aggregate better, also exhibited increased co-aggregation when combined with cells of the parental Δ*rib4:rib5* yeast. After 5 h had passed, the strain C2 was aggregating by 5.5% more than the parental strain, but the percentage of aggregation for strain B4 at the same time point was 4.2% less than the parental strain. Therefore, as in the case of auto-aggregation it seems that there is no clear correlation between increased production of riboflavin and the ability to form co-aggregates. The most important observation from these two experiments is that the parental strain aggregates by 32% after 24 h and this percentage increases to almost 46% when the same cells are mixed together with yeast cells (Figure 2.22B). These percentages suggest that already from the start of the evolution the members of our community were able to form aggregates with each other and to successfully exchange metabolites and hence there was no evolutionary pressure acting towards the increase of aggregation. The SNPs that we detected as related to aggregation, might either not affect the activity of the protein, or might be related to different physiochemical characteristics of the bacterial cell's surface.

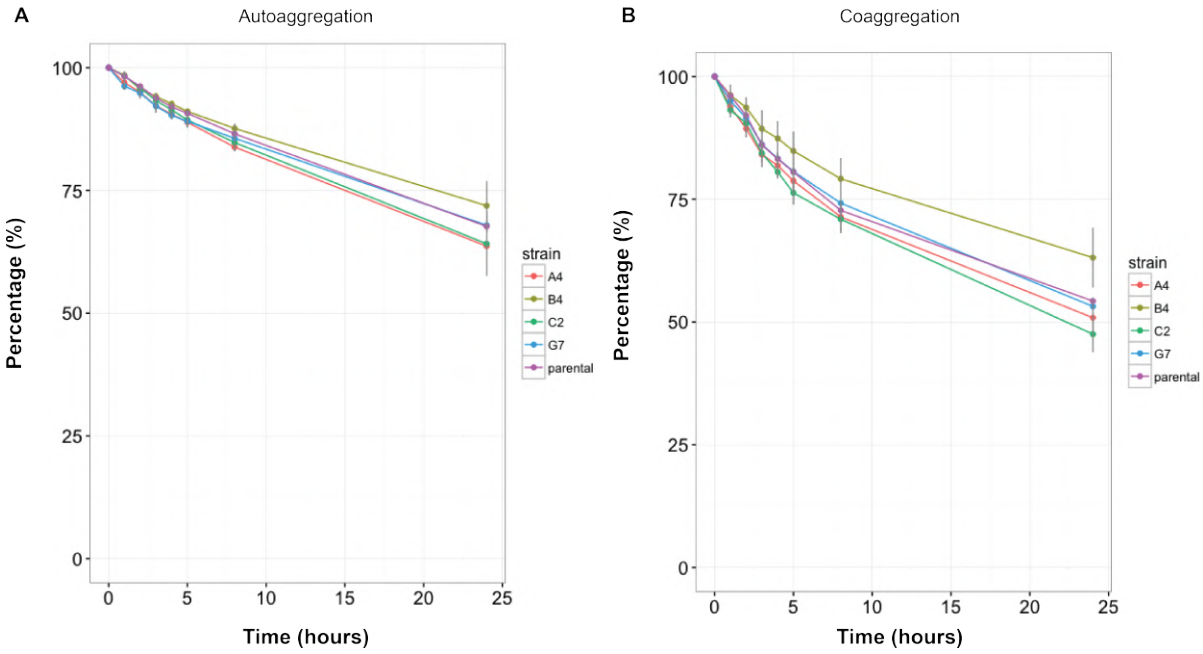

Figure 2.22. Comparison of the ability to form aggregates between the parental and four evolved strains. (A) Bacterial cells in aqueous suspension (PBS) have the ability to attach to each other over time (autoaggregation) and they exhibit a similar phenotype, (B) when they are mixed together with yeast cells (co-aggregation).

2.2.4.2 Biofilm forming capability as a potential factor that promotes interaction inside the community

Following the aggregation experiments, we tested the hydrophobicity of the cell surface of the parental and the evolved strains. The way to test with characteristic of the cell surface is by measuring the microbial adhesion to hydrocarbons (MATH), as it has been described previously (Kotzamanidis *et al.*, 2010). The differences we detected between strains were not strong, with the majority of evolved strains being less hydrophobic than their parental. In more detail strains B4, C2 and E6 exhibited a decrease of 1.4, 4 and 8.2% respectively. The two strains that were able to auto-aggregate the most, when compared to the parental were also slightly more hydrophobic, as we noticed an increase between 3.3 and 3.7% (Figure 2.23A). The increase of hydrophobicity of some strains could be associated with higher structure within the population where some strains were part of bigger aggregates or were forming biofilm.

The last experiment we conducted to test whether we could observe increased structure, was to measure the biofilm forming capabilities of the bacterial strains, with or without yeast cells being present. In our experimental setting, we measured higher values of biofilm formation for all evolved bacterial strains, independently of whether the bacteria were cultured alone or with yeasts. Nonetheless, the observed differences were higher in bacteria monocultures, rather than in co-cultures of bacteria and yeasts, particularly for strain B4 we detected a 0.72 log2fold increase in monoculture (Figure 2.23B), but in co-culture we observed a 0.11 log2fold increase compared to the parental. On the other hand, culture conditions didn’t affect strain C2, as the log2fold increase compared to the parental in monoculture was equal to 0.35 and in co-culture was 0.36 (Figure 2.23C). The fact that all evolved strains are able to form biofilm more efficiently than the parental can hint towards the importance of a specific structure for the evolution of the community, but we need to keep in mind that the conditions under which we chose to perform the ALE also favour the formation of biofilm.

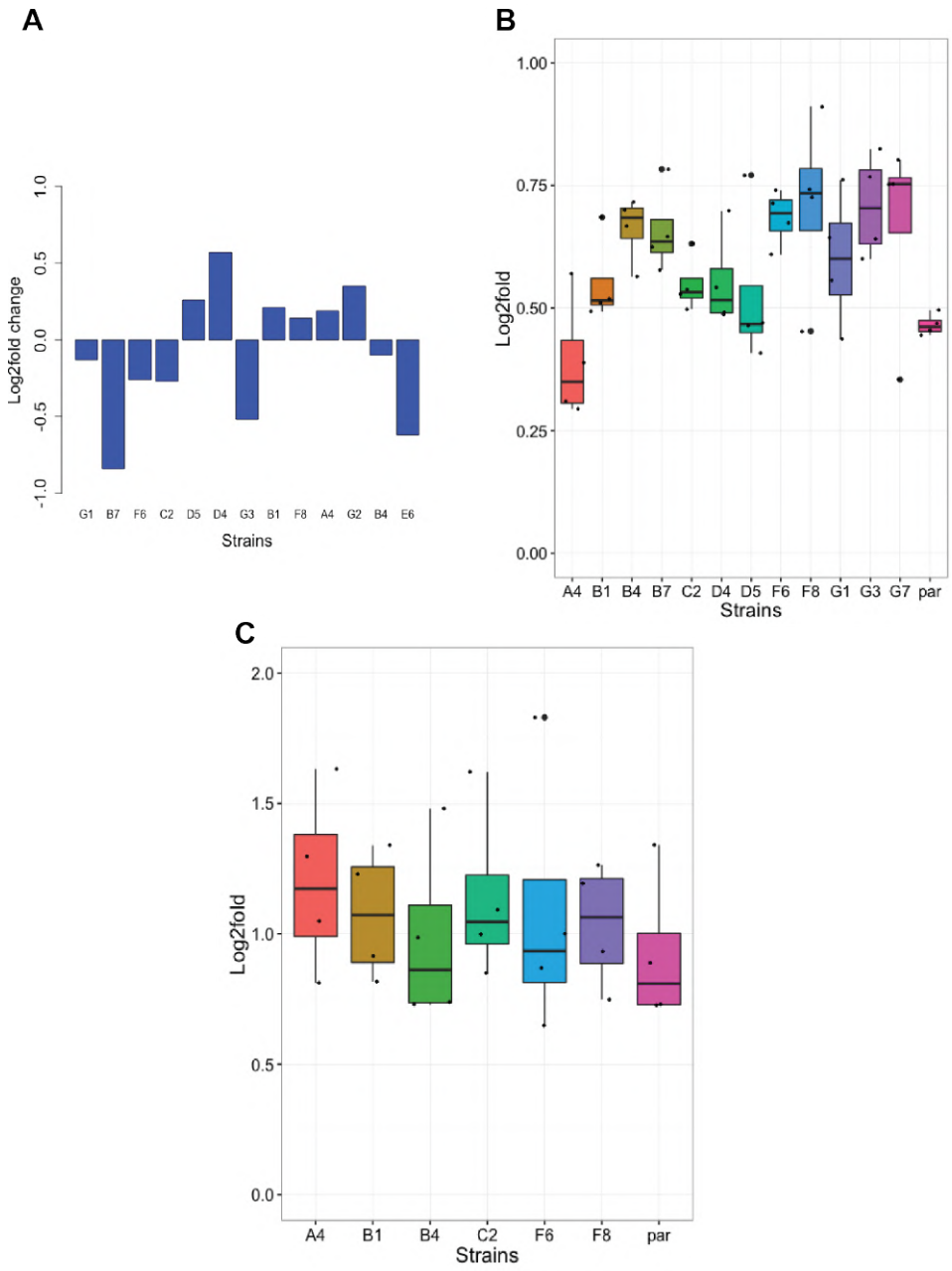

Figure 2.23. Evaluation of hydrophobicity and biofilm forming ability. (A) Comparison (log_2fold change) of the ability of evolved strains and their parental to adhere to hydrocarbons (MATH). This test is used to determine the hydrophobicity of the cell surface of bacteria. Estimation of the difference (log_2fold change) between the parental strain and evolved strain to form biofilm, either (B) as monoculture, or (C) in co-culture with the parental auxotrophic yeast. Each replicate is represented by a black dot.

2.2.5 Metabolomics analysis sheds light on the phenotypic changes

To better understand how the community evolution led to the observed differences in the phenotype, we looked into the differences that occurred during the evolution process at the metabolism level. Riboflavin production is mostly connected to glycolysis, the pentose phosphate pathway and the metabolism of nucleotides and thus we looked for differences in these pathways. Moreover, we decided to look for potential differences in the uptake and metabolism of amino acids, which play an important role in our system of community evolution. Detection of a metabolite through the available analytical methodologies depends on the physiochemical characteristics of each compound, such as the size and polarity of the molecule. For this reason, in the first part of this chapter, I will describe the detection of nucleotides and their related compounds by LCMS. All the other metabolites that are described and that are associated with the metabolism of amino acids, glycolysis, the pentose phosphate pathway and the TCA cycle were detected by gas chromatography-mass spectrometry (GCMS) with a method developed in our group (Strucko *et al.*, 2018). Similarly to the proteome analysis chapter we select as significant metabolites that exhibit a difference of at least 1 or -1 log_2fold change (p-value < 0.05).

2.2.5.1 Comparison of the intracellular pools of nucleotides

The differences we observed in the pools of nucleotide precursors were not very significant between evolved strains and their parental (Figure 2.24). Strain E6 had a 1.1-log_2fold increase in available GMP, which can be associated with the importance of GTP as a precursor of riboflavin. On the other hand, we did not detect significantly increased amounts of inosine monophosphate (IMP), which is an important intermediate in purine metabolism. The main difference detected was for strain C2 (Table 2.1) with a 0.5-log_2fold increase, while strains E6 and G7 (Table 2.1) exhibited relatively similar amounts as the parental. Changes in the levels of AMP were similar to the changes described for IMP, but still were not significant. Finally, all the tested strains had higher levels of UMP ranging from an average of 1.05-log_2fold change for strain C2 to 1.6-log_2fold change for strain E6.

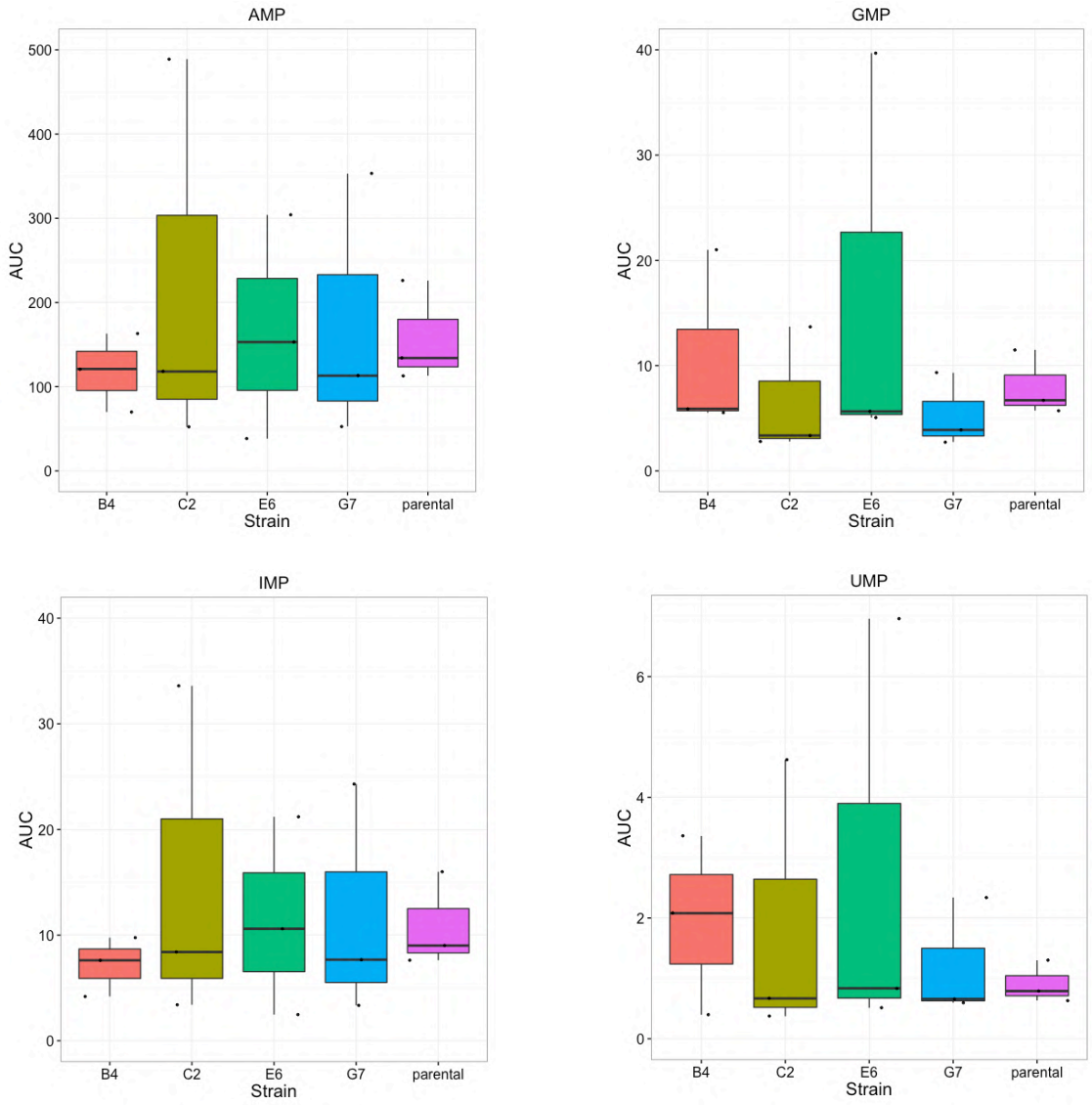

Figure 2.24. LCMS measurement of the intracellular concentrations of nucleotide monophosphates. The measurements were performed for samples of the parental and three evolved strains. In each panel, the area under the curve (AUC) of each nucleotide precursor (left to right: AMP, GMP, IMP and UMP) is depicted with each of the replicates being represented by a black dot.

2.2.5.2 The extracellular metabolism remains largely intact

Next, we analysed extracellular and intracellular samples of bacterial monocultures in CDM46R- with GCMS and compared the metabolic profile of the evolved strains with the parental strain. The analysis of extracellular samples showed that all the evolved strains maintain the same metabolic activity as the parental strain, at least concerning their fermentation profile and central carbon metabolism. First of all, there were not significant changes in the levels of lactic acid and the only detected differences that can be associated with carbon metabolism are the higher amount of succinate detected in strains C2 and G7 (1.3 and 1.5 $\log_2$fold increase, respectively) and the decrease of glycerol in strain E6 (–2.1 $\log_2$fold change). The only detected difference in the amount of amino acids was observed

in the consumption of histidine by strain E6, as we detected a decrease of 4.0 log_2fold in the medium, which can be associated with increased riboflavin production since this specific amino acid can be converted into glutamate, an important precursor of *de novo* purine biosynthesis, as it will be described with more detail in the next paragraph (Fernández and Zúñiga, 2006). A possible explanation for the limited number of observed differences related to nitrogen metabolism might be that for this experiment we used fully supplemented CDM46R-, which contains every amino acid in high concentration and thus detecting subtle differences between the strains was not possible. Moreover, it is possible that at the growth stage, in which the cells were collected, the concentration of amino acids had not decreased enough yet to allow an accurate estimation of the difference between the strains.

2.2.5.3 Investigation of differences in the Intracellular metabolism of amino acids

On the other hand, we were able to detect differences between the strains, when we analysed their intracellular metabolites. From previous work in our group we knew that threonine, glutamine, alanine, glutamate, serine and glycine were identified in the set of amino acids being secreted by the wild-type S90 yeast strain (Ponomarova *et al.*, 2017). All evolved strains exhibited higher concentration for the majority of these amino acids (Figure 2.25). In more detail, we detected increased amounts of glycine and threonine in all the samples originating from the evolved strains. Specifically, the log_2fold change of glycine ranged from 1.5 for strains B4 and G7 to 1.9 and 2.2 for the strains C2 and E6 respectively. The corresponding increased values for threonine were ranging between 1.2 and 1.4 log_2fold change for all the strains. These two amino acids are of particular interest, since glycine is an important precursor of purine biosynthesis, while the main catabolic pathway of threonine in LAB involves the conversion of this amino acid into glycine and acetaldehyde, which is a major flavour compound of yoghurt. Evidence of the role of glycine in the biosynthesis of riboflavin can be clearly found in the metabolism of the fungus *Ashbya gossypii*. When cultures of this natural riboflavin overproducing species were supplemented with glycine, riboflavin production was increased by 30% (Schlüpen *et al*., 2003) and for the same fungus it has also been shown that this amino acid is incorporated into the riboflavin molecule (Plaut, 1954). However, glycine metabolism in LAB remains largely unknown, with the best described enzymatic reaction in *L. plantarum* so far being the reversible transfer of the amino group from glycine to α-oxoglutarate. Therefore, the increased preference for catabolism of glycine and threonine by our evolved strains can be associated with increased availability of necessary precursors for the production of riboflavin. This observation can be supported further by taking into consideration the difference in the levels of glycine and threonine

between strains. The parental strain had 1.9 times more glycine than threonine, while the evolved strains had up to 3.3 times more glycine than threonine, suggesting increased catabolism of threonine.

A small increase of around 1.3 log_2fold change was also observed for serine, in strains C2, E6 and G7. In LAB, serine can be either converted directly into pyruvate, or reversibly to glycine and tetrahydrofolate. The parental strain exhibited increased pools of serine compared to glycine, as serine was 1.8 times more than glycine with the same value for the evolved strains being closer to 1 (1.1 to 1.4) and especially for E6 where it was 0.9. This observation can be associated with less conversion of glycine to serine or higher consumption of serine for the production of pyruvate. Furthermore, the levels of glutamine were found considerably increased in strains C2, E6 and G7, with an average log_2fold change of 5, a fact that can be attributed to the role that glutamine plays in the *de novo* biosynthesis of purines, where it reacts with phosphoribosyl pyrophosphate (PRPP) to form IMP which in turn is transformed into AMP and GMP (Shi *et al*., 2014).

It is worth mentioning that there was no significant change observed between the strains for alanine, which is another amino acid secreted by the auxotrophic yeast, but its accumulation seems not to be favoured by the bacteria. It has been described that some strains of LAB can metabolise alanine when α-oxoglutarate is available, or alternatively alanine can be degraded to pyruvate and ammonia via a D-amino dehydrogenase, or be converted into pyruvate by transamination, however, neither of these reactions have been described so far in *L. plantarum*. Another amino acid for which we did not observe differences was cysteine, which can be either degraded to ammonia, hydrogen sulphide and pyruvate, or can be utilised for the biosynthesis of methionine. However, as in the case of alanine, it seems that the use of cysteine is a strain related trait (Fernández and Zúñiga, 2006) and that our strains prefer serine for the synthesis of pyruvate. The last example of non-preferred amino acid is aspartic acid, which is closely related to alanine and can potentially be metabolised to fumarate and pyruvate. Nevertheless, a larger number of enzymatic steps are required for these reactions compared to the metabolism of serine and perhaps this is the reason why the evolved strains prefer to consume the latter. Finally, we detected much higher levels of ornithine in three of the evolved strains, ranging from 3 to 5 log_2fold difference compared to the parental strain, which is a hint that during the evolution bacteria evolved to better regulate the intracellular amount of nitrogen, to avoid poisoning from excess concentration. These examples enhance the hypothesis we formed that throughout the evolution process the bacterial cells adapted

towards the preferred use of available nitrogen sources, increasing their bias towards the use of amino acids that can provide the necessary precursors for the synthesis of riboflavin.

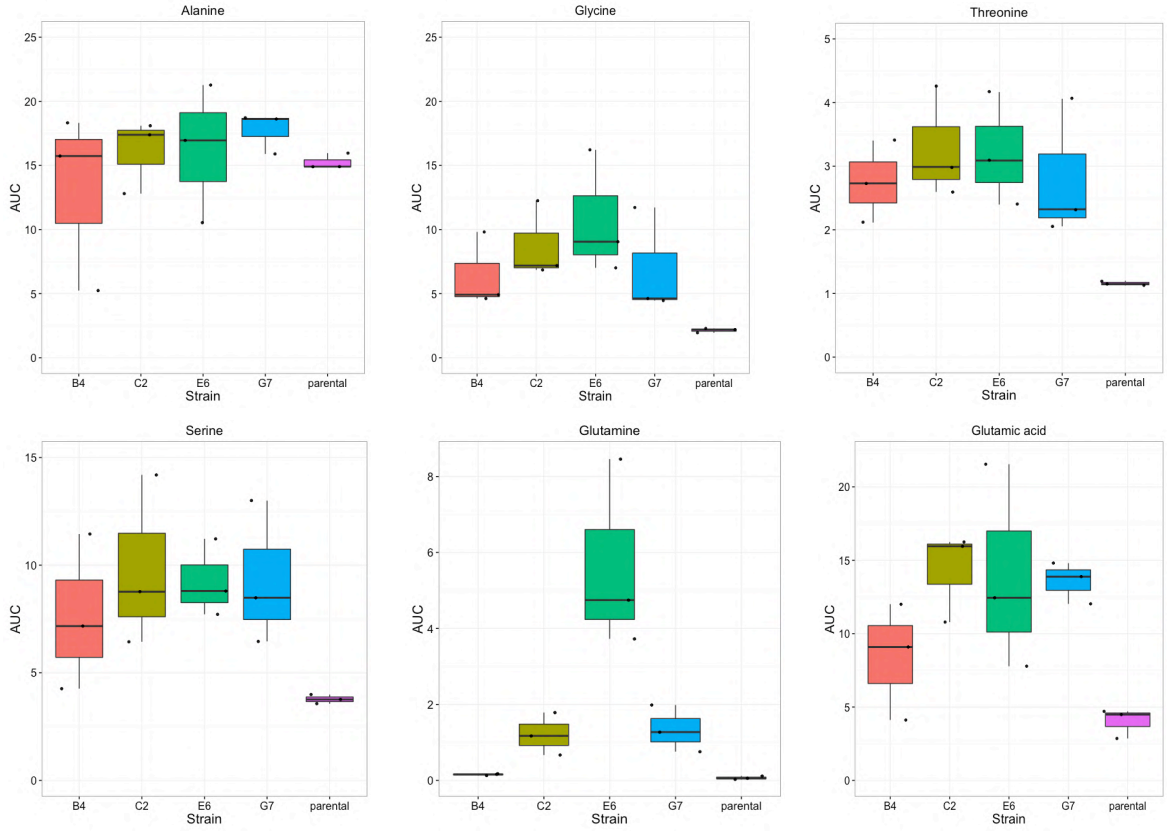

Figure 2.25. GCMS measurement of the intracellular concentrations of amino acids. The samples were collected from bacterial monocultures in CDMD46R-. The depicted amino acids are described as being secreted by *S. cerevisiae* under co-culture conditions. In each panel the area under the curve (AUC) of each amino acid (left to right: alanine, glycine, threonine, serine, glutamine and glutamic acid) is depicted with each of the replicates being represented by a black dot.

2.2.5.4 Examination of the central carbon metabolism

Apart from differences related to nitrogen metabolism we could also identify similarities and differences connected to glycolysis and the pentose phosphate pathway. Lactic acid is one of the most important compounds of LAB metabolism and is largely used to describe this clade of bacteria. All our strains exhibited similar levels of intracellular lactic acid (Figure 2.26), an observation which agrees with the fact that also the levels of extracellular lactic acid were also very similar. Moreover, we did not detect significant differences in the levels of pyruvate and its direct precursor phosphoenolypyruvate (PEP). Together, these results lead us to suggest that glycolysis remained largely unchanged through the evolution process. At the same time, all the evolved strains have higher pools of tricarboxylic acid cycle intermediates and more specifically malate (2.2 $\log_2$fold change) and fumarate (1.6 $\log_2$fold change) in the case of strain B4, fumarate (2.1 $\log_2$fold change) and succinate (1.7 $\log_2$fold

change) for strain E6. Both strains C2 and G7 had higher amounts of all three previously mentioned metabolites (Figure 2.26). However, the observed increase in citric acid cycle intermediates cannot be associated in a straightforward manner with increased riboflavin production.

The pentose phosphate pathway (PPP) plays an important role in the biosynthetic pathway of riboflavin as the important precursor ribulose 5-phospate is one of the main products of the PPP, which starts with the dehydrogenation of glucose 6-phospate (G6P). We were able to detect higher values of G6P for all the evolved strains, with high riboflavin producing strains B4 and E6 exhibiting a $\log_2$fold change of 1 compared to the parental and for strains C2 and G7 the same difference was equal to 1.5. Moreover, we could not detect statistically significant differences between the samples from the parental and evolved strains for the metabolite ribose-5-phosphate, which is formed through the enzymatic conversion of ribulose 5-phosphate in the PPP. The similarity in the detected amounts of lactic acid and ribose-5-phosphate, leads us to assume that the increased amount of glucose 6-phospate was not diverted solely to glycolysis and that the potentially increased amounts of ribulose 5-phosphate are fed into the synthetic pathway of riboflavin. However, some amount of ribulose 5-phosphate is turned into ribose-5-phosphate which is further metabolised into PRPP for the biosynthesis of both purines and pyrimidines.

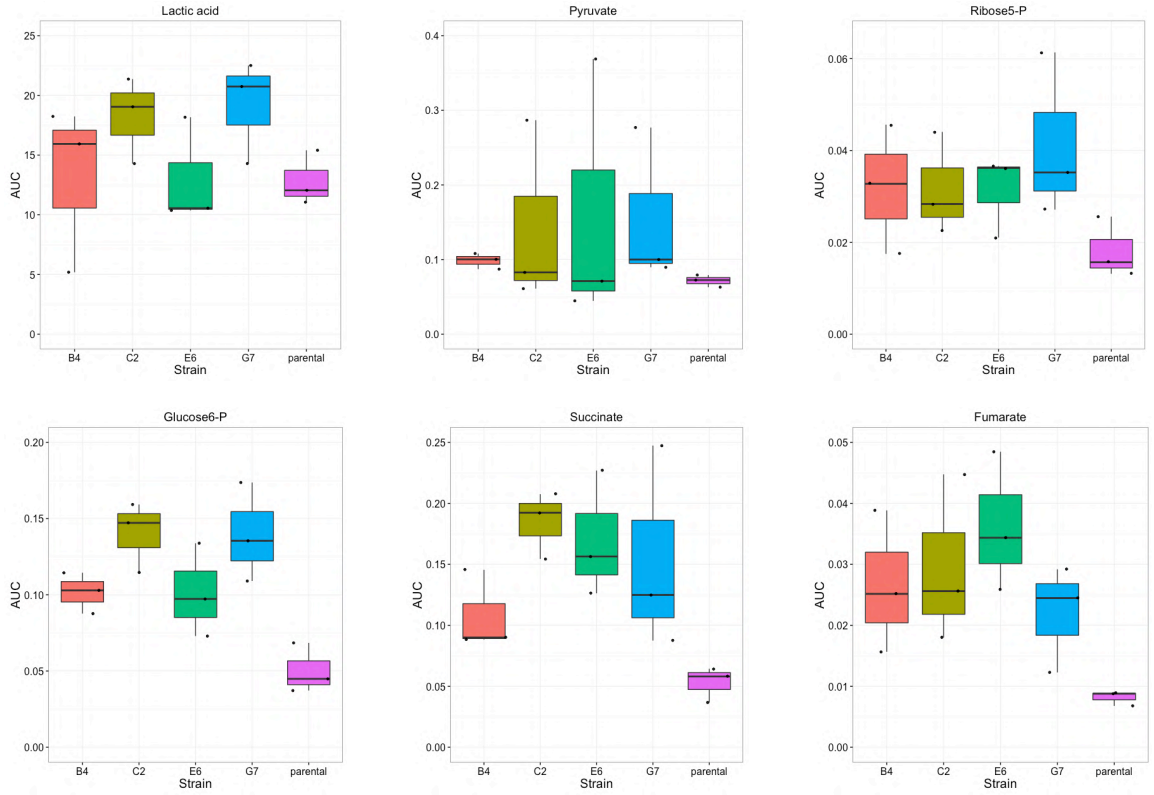

Figure 2.26. GCMS measurement of the intracellular concentrations of metabolites related to carbon metabolism. The samples were collected from bacterial monocultures in CDMD46R-. The depicted metabolites are associated with either glycolysis, the PPP or the TCA cycle. In each panel, the area under the curve (AUC) of one metabolite (left to right: lactic acid, pyruvate, ribose-5P, glucose-6P, succinate and fumarate) is depicted with each of the replicates being represented with a black dot.

As previously described, the differences between the evolved and the parental strains concerning nucleotide precursors were not significant, nevertheless, uracil and its precursor UMP were more abundant in the samples originating from 3 of the 4 evolved strains, with strain B4 being the only one with similar values as the parental strain. The difference for the other strains was more prominent for strains C2 and G7 with a difference of 2.4 and 2.7 log_2fold respectively, when the log_2fold change for strain E6 equal to 2. This observation seems puzzling, as conversion of uracil to UMP requires PRPP, which is also a necessary precursor of the purine biosynthetic pathway. This pathway leads to the formation of GTP, which then can be used for riboflavin production. A potential explanation about this observation can be found in literature and more specifically in research conducted on the riboflavin overproducing fungus *Ashbya gossypii*. In *A. gossypii* increased levels of available PRPP have been associated with riboflavin overproducing phenotypes, especially in pyrimidine auxotrophic mutants (Silva *et al.*, 2015). It has been suggested that in *A. gossypii* high concentrations of uracil negatively regulates the activity of AgUPRT, an enzyme with uracil phosphoribosyltransferase activity, which participates in the pyrimidine salvage pathway. This enzyme has PRPP and uracil binding regions and hence decreased activity can

lead to higher availability of PRPP for the biosynthesis of purines (Silva *et al*., 2019). Something similar has not been suggested for the well-studied *Bacillus subtilis*, which is the preferred industrial riboflavin overproducing bacteria species, but our metabolomics analysis suggests that a similar mechanism might be present in bacteria as well.

2.2.5.5 Summary

The overview of all the described observations from the intracellular samples are depicted as a heatmap, to better visualise the similarities and the differences between the evolved strains, in comparison to their parental strain (Figure 2.27). In short, analysis of the extracellular metabolism, as well as of the intracellular nucleotide pools didn't highlight major differences between the parental and the evolved strains. Therefore, we were not able to confirm the previously descripted upregulation of enzymes that participate in the biosynthetic pathway of purines, as it was descripted in the chapter about the proteomics analysis. On the other hand, through the analysis of the intracellular metabolites we were able to detect increased amounts of compounds that participate in the riboflavin biosynthetic pathway, such as ribose-5-phosphate. Finally, analysis of the intracellular concentration of amino acids showed that the levels of some amino acids, such as alanine remained similar between the parental and the evolved strains, while the concentration of amino acids that participate in the biosynthesis of vitamins, such as glutamine and glutamate, were increased in the samples originating from evolved bacteria.

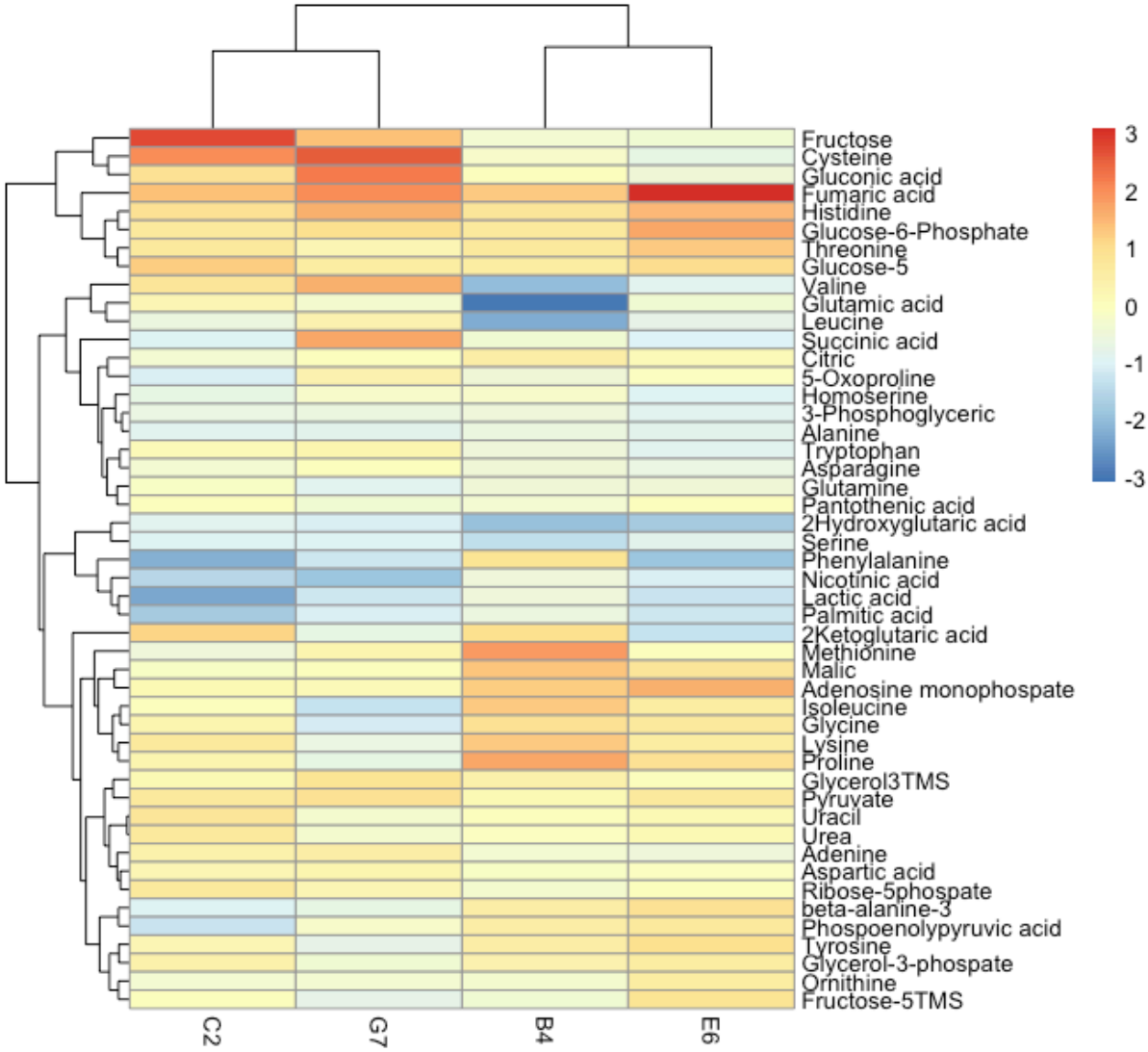

Figure 2.27. Cluster of major differentially regulated metabolites. Heatmap of the log_2fold change between the parental and the evolved strains for all the metabolites that were detected intracellularly with GCMS. The colour gradient represents the value of the calculated log_2fold change.

When we compare the two different heatmaps (Figures 2.19 and 2.27) we observe that the strains that we characterised as the best riboflavin producers, i.e. B4 and E6 (Tables 2.1 and 2.2) cluster together, independently of whether we compare their proteome or their metabolism. In a similar fashion, strains C2 and G7 which did not express such a strong riboflavin secreting phenotype (Table 2.1) form in both cases a separate cluster together.

2.2.6 Adaptive laboratory evolution for increased production of folate

2.2.6.1 Establishing a metabolically dependent community for folate overproduction

Since community evolution led to the improved production of riboflavin, we moved on to showcase that the same principle can be applied for the production of the second compound we had initially chosen for the present study, meaning folate (vitamin B9). The parental *L. plantarum* strain exhibits also the ability to produce this member of the B vitamins group. Microbial production of folate is not straightforward, as one molecule of folate contains one pterin moiety bound to para-aminobenzoic acid (pABA), which is produced from the pentose phosphate pathway. However, genome analysis of lactic acid bacteria showed that many strains that contained the genes responsible for folate biosynthesis were not able to synthesise pABA and their phenotype was auxotrophic for folate (Rossi *et al.*, 2011). Like in the case of riboflavin, *S. cerevisiae* is not dependent on environmental folate, as it is able to synthesise both pABA and folate. At the same time if the pathway for folate biosynthesis is disrupted, but pABA synthesis is undisrupted, the cells can recycle other compounds like nucleotides, to produce folate (Rossi *et al.*, 2011). For this reason, we decided to engineer the auxotrophic phenotype to the parental yeast strain (abz1::kanMX4) and use it with the same parental *L. plantarum* GKC strain that was used for increased riboflavin production. The evolution environment was a modified version of CDM35 medium, where both pABA and folate were omitted and riboflavin was added (named CDM33F-). Similarly, to the process that we followed for increased production of riboflavin, initially we tested whether the two strains can form a stable community. When we were culturing them together we could detect sufficient growth, therefore we could use this community to perform a directed evolution experiment. As in the case of riboflavin we started the experiment with twelve identical populations, which were evolved in parallel.

2.2.6.2 Laboratory evolution and determination of phenotypic changes

During the evolution period the optical density of the twelve populations increased between 4 and 6 times, with visible oscillations between the transfers, similarly to what was observed during the evolution of riboflavin producing communities. In contrast to these communities though, the evolution process was slower, as the 5 first transfers were done every 5 days and after that the time between transfers was reduced to 3 days (Figure 2.28A). Also, in this experiment the time between transfers was adjusted based on the required period for complete growth of all the replicates. After the completion of the serial transfers single colonies of *L. plantarum* were isolated for further characterisation from the final populations.

In contrast to the riboflavin community evolution for riboflavin production we could not use a sensitive enough screening method to initially pick strains that exhibited an improved phenotype. For this reason, we decided to directly test folate production of a small number of isolates by UPLC (Figure 2.28B). In total, the secreted amount of folate was determined for four evolved strains and their parental. As in the case of riboflavin we chose to measure the secreted concentration of direct precursors such as dihydrofolate (DHF) and tetrahydrofolate (THF). In the UPLC method that we chose, the commercially purchased standards of both DHF and THF eluted as 3 distinct peaks with the same retention times, while pure folate behaved similarly and eluted as two separate peaks. The two folate peaks had the same retention time as two of the three peaks that were identified previously for DHF and THF. It is worth mentioning that the THF standard is commercially available only with 65% purity, which can partially explain the elution of multiple peaks. For this reason, we took into account all the peaks that we were able to detect in the supernatant samples of our strains and eluted at the same retention time as our standards. In all the samples of the parental strain we could identify solely a single peak, which was present in all 3 of the standards and was eluting last, while the rest were either not present or below the detection limit. Apart from this same peak, we identified and measured in the samples of all the tested evolved strains one of the peaks that was present only in the DHF and THF standards. Nevertheless, one of the evolved isolates' samples had no detectable third peak, which is the peak that is common between parental and evolved strains. Therefore, it seems that like in the case of increased riboflavin production, the bacterial cells evolved to support faster growth of yeast cells by secreting increased amounts of valuable precursors, apart from the target compounds. Or it is simply a way to avoid the negative feedback loop by accumulating smaller amounts of all the related compounds inside the cell. Next, we used the peak that was common between the majority of the evolved strains and the parental to calculate the difference between them. The evolved strains were able to secrete higher amounts of the compound that behaves similarly to folate as we detected for strains A3 and C9 a 0.8 $\log_2$fold increase. The best performing strain was strain B8, for which we detected a 1.8 $\log_2$fold change compared to the parental.

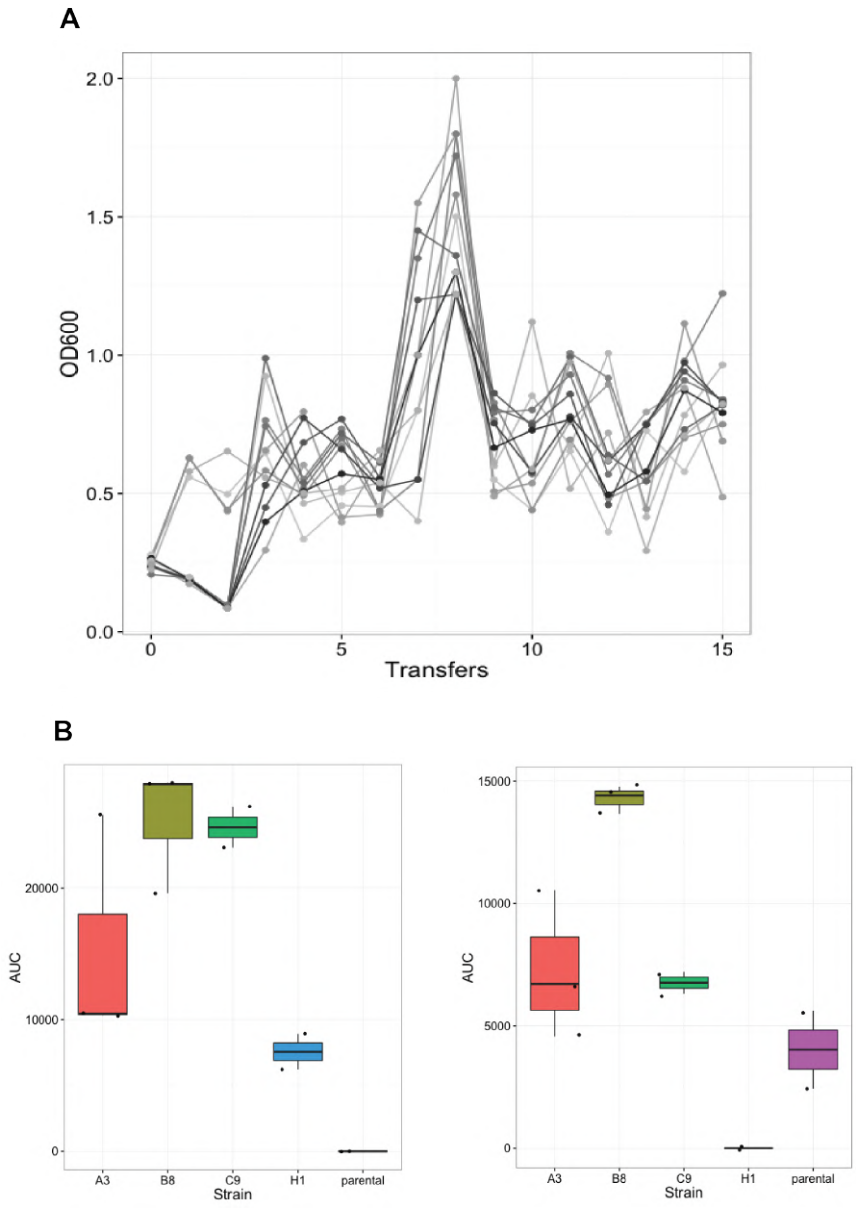

Figure 2.28. Experimental coevolution of mutualistic community for increased production of folate. (A) Co-culture of a small mutualistic community for increased folate production consisting of *L. plantarum* strain GCK and *Δabz1*. Twelve identical populations were evolved in parallel in 96-well plates. The growth of each community improved similarly for every replicate, based on optical density measurements. (B) UPLC analysis of extracellular samples from the parental *L. plantarum* strain and four evolved strains (A3, B8, C9 and H1). The two panels represent the area under the curve of the two peaks that were identified as THF and folate according to commercially purchased standards. Each value of the replicates is represented with a black dot.

Based on the improvement of the OD_{600} values during the ALE co-culture and the UPLC analysis, we can assume that the evolution process for increased production was successful. However, in this case we could not use a fast screening methods, as we did previously with use of fluorescence and therefore we cannot estimate accurately which percentage of the population that improved its phenotype. Preliminary experiments where we used a biological assay, exhibited low sensitivity (data not shown) and as a result we

didn't pursue to continue using this assay. Instead we moved on to characterise further the isolated strains that exhibited an improved phenotype through UPLC.

2.2.6.3 Estimation of the relative fitness of the evolved strains

Similar to the steps we followed in the case of riboflavin, we compared the growth of the parental and the evolved strains in CDM45F- (Figure 2.29). Unlike the riboflavin over-producing strains, strains evolved for folate production were able to grow to a higher maximum OD_{600} than their parental and their growth rate was higher on average as well. For example, strain B8 which exhibited the best phenotype according to the UPLC analysis, had the same growth rate as the parental, but the final max OD_{600} was increased by 17%. Interestingly strain C9 which was the second-best producing strain exhibited an increase of 25% in growth rate as well as a 22% increase in max OD_{600}. Therefore, it seems that the production of folate is more challenging than the production of riboflavin at least for the parental strain when it was introduced in CDM33F- for the first time, either with amino acid supplementation or in mutualistic co-culture, as the evolution process for folate was slower.

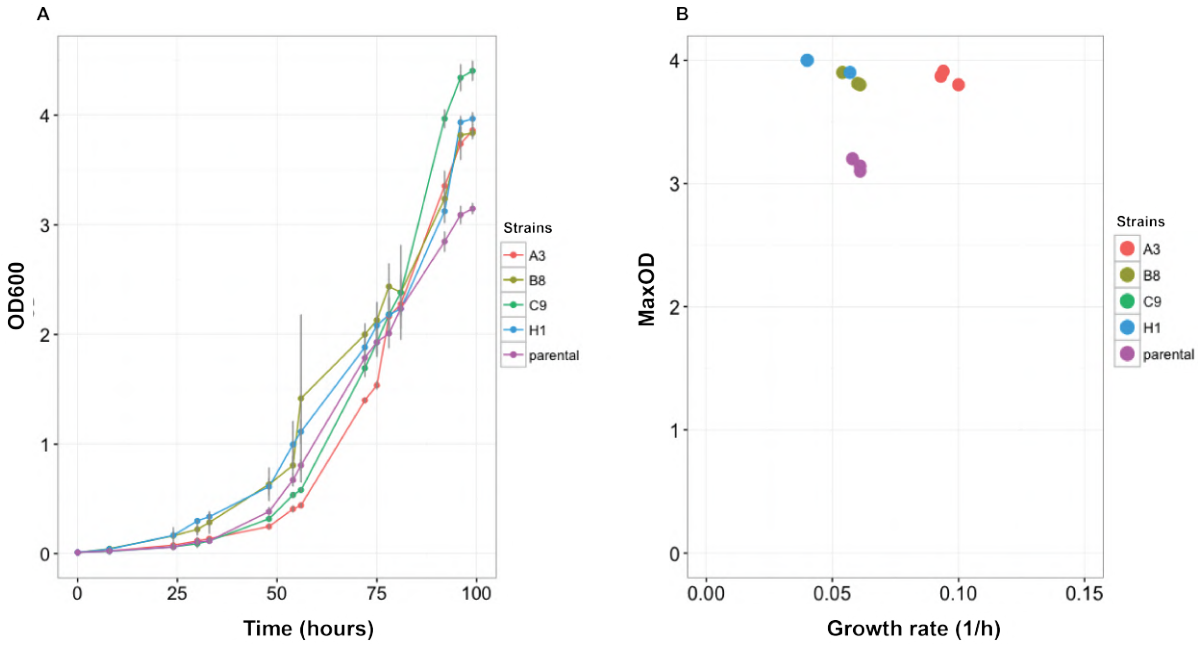

Figure 2.29. Comparison of relative fitness of bacterial monocultures. (A) Growth kinetics of the parental strain and four evolved strains in CDM46R-. In the figure the average values of biological triplicates are represented. (B) Scatter plot representing the connection between growth rate and maximal OD_{600} for each of the replicates from panel A.

Next, we set out to investigate whether conditioned medium of these newly evolved strains has the same effect on yeast growth as we previously observed for the strains evolved for increased riboflavin production. As for riboflavin, conditioned medium of these strains

can support the growth of Δ*abz1* adequately, however, the differences between the parental strain and the evolved ones were not very prominent (Figure 2.30A). All yeast cultures independently of whether they were grown in conditioned medium from the parental *L. plantarum* or an evolved strain, exhibited similar growth rates, while the maximum OD_{600} of cultures grown in spent media of evolved strains was increased by 15%, compared to the ones grown in conditioned medium of the parental strain. The emergence of only minor differences might be associated with generally lower required amounts of folate that can restore the phenotype of the auxotrophic yeast. On the other hand, the difference in growth was more prominent between parental and evolved bacteria cells when they were grown in conditioned medium of wild-type S90 yeast (Figure 2.30B). In more detail, evolved isolates needed about 24 h before they exhibited growth in the conditioned medium and after 48 h they had reached the highest recorded values of OD_{600}. These values were about 1.3 times (30%) higher than the same value that we recorded for the parental strain. In conclusion, it seems that the strains evolved for the production of folate developed similar adaptations as the strains evolved for the production of riboflavin, meaning an increased ability to uptake amino acids secreted by their yeast partner, while at the same time they were able to support growth of the yeast cells more efficiently. Finally, we decided to sequence these isolates, in hope of identifying mutations that could be potentially linked to the observed phenotypic improvement.

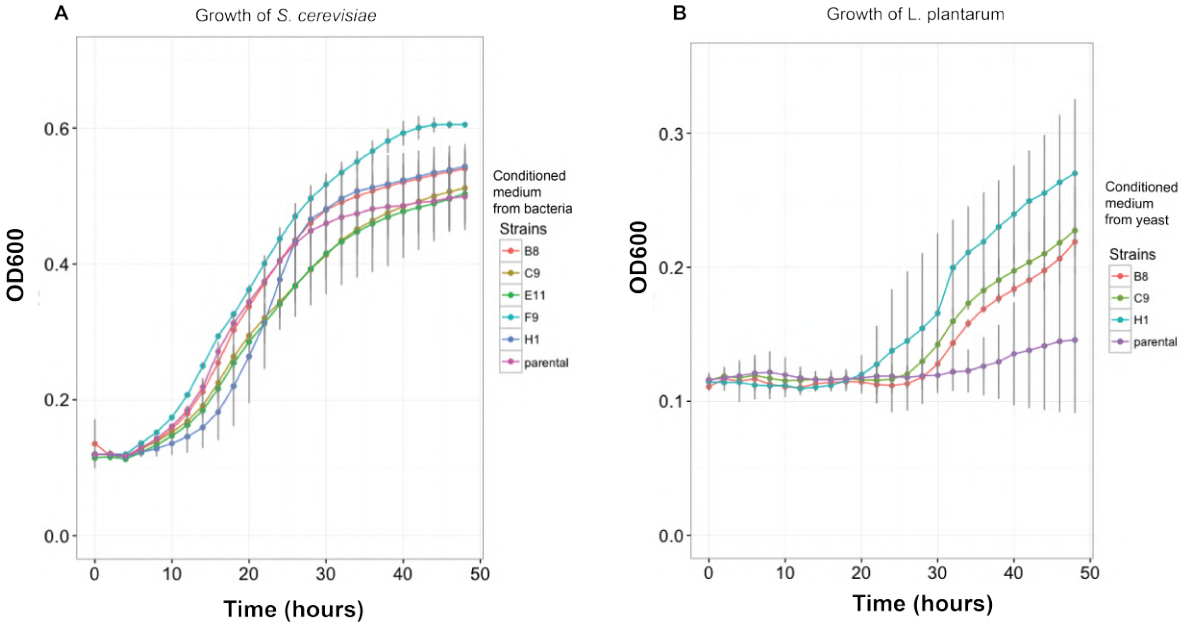

Figure 2.2.30. Comparison of relative fitness in conditioned media. (A) Growth kinetics of the parental *Δabz1 S. cerevisiae* strain in conditioned medium of the parental and evolved bacteria strains. In the figure the average value of biological replicates is represented. (B) Growth kinetics of the parental and evolved bacteria strains in conditioned medium from wild-type *S. cerevisiae*. In the figure, the average values of biological triplicates are represented.

2.2.6.4 Whole genome sequencing analysis reveals similarities between the two co-cultures

Similarly, to the procedure that we followed for the riboflavin evolution, we sequenced the clones that exhibited an improved folate producing phenotype, aiming to identify potential causative mutations. As in the case of riboflavin we could not find direct genomic evidence linked to the production of folate. It seems that the selected evolutionary pressure led the strains to adopt a common strategy of adaptation due to the fact that we identified some shared mutations between the two different experiments, and more prominent mutations on amino acid transporters (amino acid ABC transporter substrate-binding protein), cell surface proteins or proteins that interact with the extracellular environment (mucus-binding protein) and transcription regulators that belong to the same families, like the MarR transcription regulators and the TetR transcription regulators. This group of related mutations, which seem to be associated with the conditions of the coevolution experiment, includes the most important identified mutations. Nevertheless, we were able to identify some mutations that occurred solely in one of the two evolution experiments. For example, in strain C9, we identified a SNP in the coding region of the gene that encodes for diguanylate cyclase. This enzyme catalyses the formation of cyclic-di-GMP from two molecules of GTP. Like in the case of riboflavin GTP is also important for the biosynthesis of folate.

2.2.6.5. Summary

The experiments we described in this chapter are proof that use of mutualistic communities to perform ALE can be extended towards the increased production of more compounds, apart from riboflavin. Specifically, we showed that creating the correct conditions for mutualism, resulted into increased production and secretion of folate. The duration and the speed of the ALE experiment varied between the two cases, meaning that even though the general parameters remain the same, the details for the production of different compounds vary. Independently of whether the target compound was riboflavin or folate, evolution of the co-culture resulted to the emergence of bacterial strains that could better support their partner, while at the same time these strains were able to assimilate better the limited nitrogen sources that were made available to them. By sequencing the genome of evolved bacterial strains from both cases, we were able to identify some similarities, which can be attributed to the similar environmental conditions we used and to the fact that we used the same species in both cases. In more detail, both strains evolved to produce riboflavin and strains evolved to produce folate accumulate point mutations in genes connected to transcriptional regulation and transport of amino acids, while it seems that mutations directly

linked on the biosynthetic pathway of each vitamin are not favoured. Scientific community is trying to establish laboratory evolution as a method with a directed result, based on the experimental design. So far use of mutualistic communities had resulted into to the acquisition of the desired outcome. Nevertheless, it was unclear whether this outcome resulted from the selection pressure created by the specificity of the metabolic cross feeding, or from stochastic events occurring during adaptation. For this reason, we set out to test our hypothesis that the phenotype can be improved only when the bacteria evolve in the presence of their yeast partner and that each coevolution leads to increased production only of the target compound.

2.2.7 Mutualism creates the appropriate conditions for directed evolution

2.2.7.1 Evolution of bacteria in monoculture with amino acid supplementation

An important aspect of this study was to evaluate whether the mutualistic community creates the appropriate pressure for the vitamin overproducing phenotype, or whether it is a phenotype that may arise stochastically in the chemically environment of the evolution niche. As described, the yeast strain used to assemble the community has the ability to secrete amino acids which can be used as precursors and feed the pathway of riboflavin production. Hence to answer the question whether the observed phenotypic differences were related to the increased availability of precursors (i.e. amino acids), we performed a control monoculture ALE experiment where the original parental *L. plantarum* strain was evolved in CDM46R- in absence of its yeast partner. Four different populations of the *L. plantarum* strain were evolved for about 100 generations. At the end of the experiment bacterial clones were isolated from selective medium and their ability to produce riboflavin was evaluated by a fluorescence intensity based assay.

In total 100 clones were tested from the newly evolved populations and some variability was identified. The majority of strains (59%), performed worse than the parental with the $\log_2$fold change ranging from -2 to -0.25 (Figure 2.31A). About 29% performed slightly better than the parental with the $\log_2$fold change varying between 0.01 and 0.25. This increase is quite smaller compared to the differences we detected during the community evolution, when we isolated strains with a tenfold higher increase than this, ranging from 0.15 to 2.5 $\log_2$fold difference. In contrast to these results and as it was described in 2.2.1.2, more than 60% (Figure 2.31B) of the isolates originating from the community evolution performed better than the parental strain. Taken together, these results indicate that the

observed over-producing phenotype is not solely related to higher availability of precursors, but is directly linked to the evolution under mutualistic conditions.

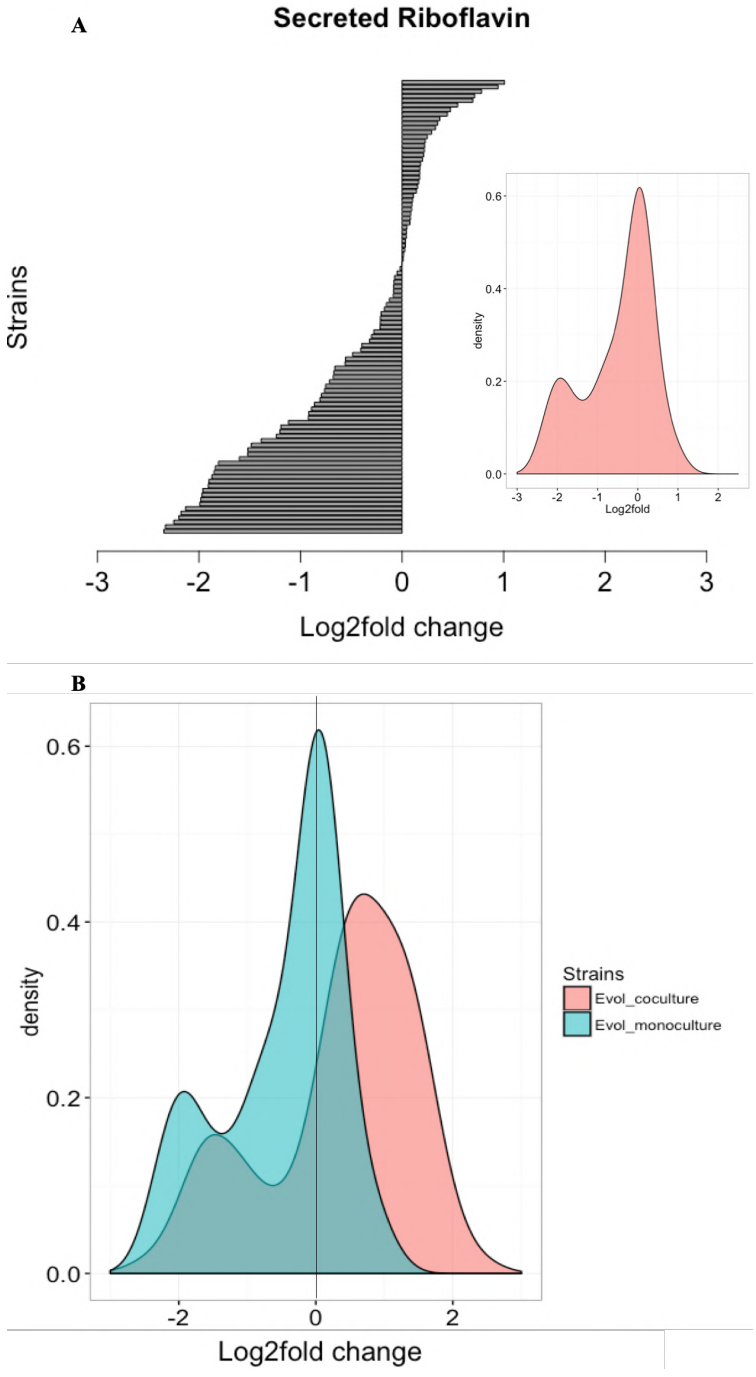

Figure 2.31. Comparison of the riboflavin producing phenotype of differently evolved strains. (A) Log_2fold change of the riboflavin concentration from supernatant of *L. plantarum* isolates monocultures, compared to the parental strain. The isolates were evolved in monoculture in fully amino acid supplemented media (CDM46R-) and the riboflavin concentration was estimated based on fluorescence measurements at 440/520 nm. (B) Density plot of all the phenotyped isolated clones, comparing the percentage of improved strains as they were isolated from the co-culture (Evol_coculture, 134 strains) and the monoculture (Evol_monoculture, 100 strains) evolution experiments. On the *x*-axis is plotted the log_2fold change of riboflavin concentration in comparison to the parental strain, as it was measured by fluorescence.

2.2.7.2 Riboflavin secretion from strains evolved for folate production

Another aspect that we needed to examine was whether the overproducing phenotype is specific for the targeted compounds, or the evolution pressure leads to an unspecific change in related pathways. For this purpose, clones that had been isolated from the evolution for increased folate production were cultured in CDM46R- and their riboflavin producing ability was evaluated through fluorescence. Compared to the parental strain, about half of the isolates (49%) presented an increase in riboflavin secretion (log_2fold change > 0.1), while the rest 51% of the isolates exhibited lower values (Figure 2.32A). To note that all the tested clones had an observed increase lower than the log_2fold value of 1, and much lower than most of the isolates evolved for riboflavin production, as described before (Figure 2.32B).

These numbers suggest that similar pathways such as those for riboflavin and folate that involve similar initial steps, GTP hydrolysis, might be affected, but the number of clones performing better and worse is quite similar. Based on all of these results we propose that ALE coevolution is highly specific, even for target compounds that share some biosynthetic steps.

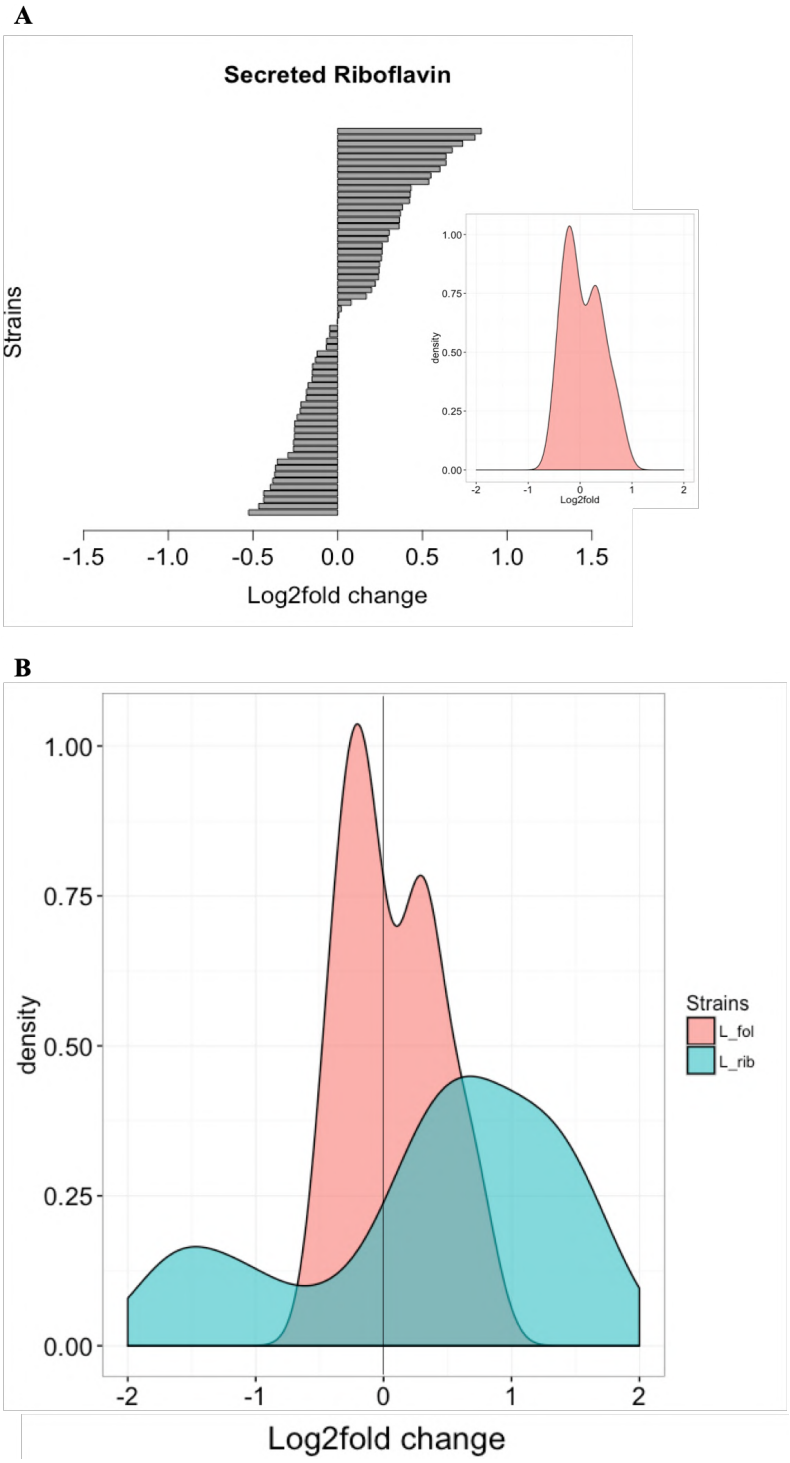

Figure 2.32. Comparison of the riboflavin producing phenotype of evolved strains originating from different ALE co-cultures. (A) Log_2fold change of the riboflavin concentration from supernatant of *L. plantarum* isolates monocultures, compared to the parental strain. The isolates were isolated from the final populations of the co-culture evolution for increased production of folate. The riboflavin concentration was estimated based on fluorescence measurements at 440/520 nm. (B) Density plot of all the phenotyped isolated clones, comparing the percentage of improved strains as they were isolated from the co-culture for production of riboflavin (L_rib, 134 strains) and the one for production of folate (L_fol, 50 strains). On the *x*-axis is plotted the log_2fold change of riboflavin concentration in comparison to the parental strain, as it was measured by fluorescence.

2.2.7.3 Summary

By evolving the parental *L. plantarum* strain in fully supplemented CDM46R- we wanted to test whether the selection pressure that acts towards the improvement of riboflavin production and secretion was the presence of the auxotrophic yeast or the lack of riboflavin from the medium. Indeed, when the bacteria are evolved as monoculture the percentage of improved strains was 3 times smaller than when they are evolved in co-culture with their yeast partner. Moreover, the maximum increase (in comparison to the parental) that was detected in clones isolated from monoculture ALE replicates was 10 times smaller than the increase described for strains evolved in co-culture. What is more, absence of riboflavin from the environment seems to favour more selfish behaviour, since the majority of the strains secrete up to 4 times less riboflavin than their parental. These observations suggest that metabolic cross feeding creates the appropriate selection pressure leading not only to increased production of the target compound but to higher secretion as well.

In similar fashion, screening of the riboflavin producing phenotype of bacterial isolates originating from the ALE experiment suggests that the selection pressure created by the mutualism is specific for each compound. Even though a few folate isolates secreted double amount of riboflavin than the parental, the majority performed either similarly or worse than the parental. The proteomics data that we acquired previously enhance the hypothesis that the outcome of the evolution was directed, in most of the samples from the tested evolved strains, protein FolE was downregulated. This protein catalyses the first step in the biosynthetic pathway of folate.

It still remained unclear though, if the outcome of the evolution process is determined by the species that form the mutualistic community. For this reason, we set out to perform similar coevolution experiments, using different species of bacteria.

2.2.8 Expansion of the coevolution approach to more species

The previously described experiments hint that our evolution system is directed towards the compound of choice, but was performed only for one specific strain of *L. plantarum* and thus we tried to perform community evolution using more species from our natural isolates collection. To better identify potential candidate strains, we reconstructed the genome-scale metabolic models for two of the species, viz. *Lactococcus lactis* and *Brevibacterium casei.* The metabolic models predicted that these two strains contain all the necessary enzymes for the complete riboflavin production pathway and are also able to excrete the compound into their environment. Moreover, since the production of folate in bacteria requires also a second co-factor (dihydropterin pyrophosphate - DHPPP), we searched in our collection for strains that can potentially produce both. The genome scale model predicted that two of the collection's *L. lactis* strains are able to produce folate without any supplementation in the medium. Each one of the selected strains was coupled with the appropriate auxotrophic yeast, Δ*rib4:rib5* for riboflavin and Δ*abz1* for folate, and serial transfer experiments started in parallel. For each one of the combinations we followed the same protocol as for *L. plantarum*. As we had already observed, communities evolved for the production of folate tend to grow slower and require longer time periods between passages, since all riboflavin communities were transferred 21 times, while at the same time folate communities were transferred 17 times. In the case of *L. lactis* which was evolved for increased production of riboflavin we observed two differently adapted communities (Figure 2.33A). About half of the replicates increased their OD_{600} by 9 times after only 2 passages, while for the rest 7 passages needed to pass before they reached the same value of optical density, while the observed oscillations between transfers were subtler than the ones we originally observed for *L. plantarum.* On the other hand, for all *B. casei* populations (Figure 2.33B) the optical density was increased almost tenfold after only 2 transfers, but experienced a sharp decline after 15 transfers with a threefold decrease.

For increased production of folate, we used two different *L. lactis* strains, which exhibited a similar pattern of adaptation during the evolution process (Figure 2.2.34). In more detail, the increase in optical density was minimal with the OD_{600} only being increased threefold by a couple of replicates in the first 7 to 8 transfers. Then after the 9^{th} transfer all the populations improved by 6 to 10 times and remained at these increased values of OD_{600} until the end of the serial transfer experiment.

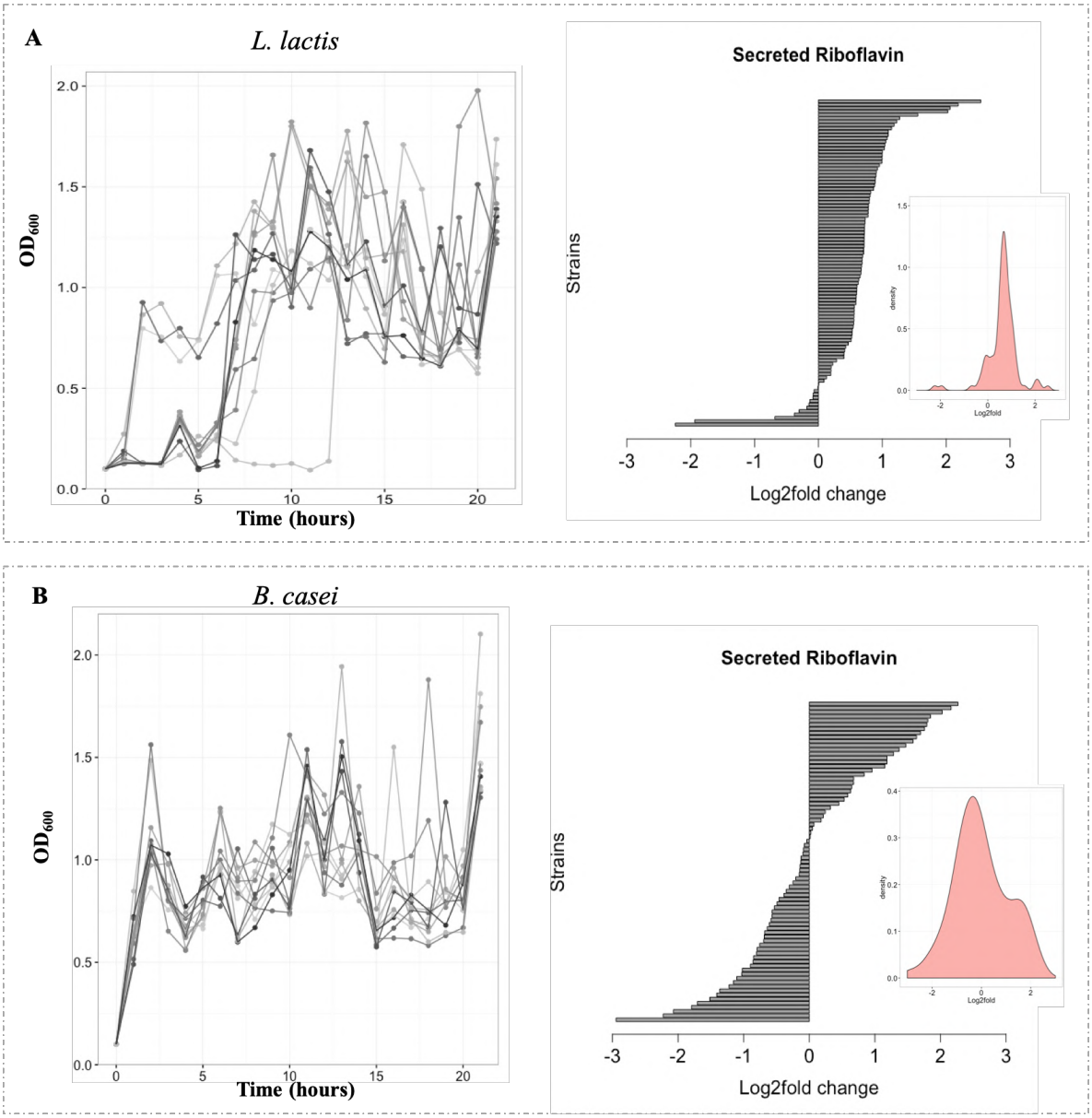

Figure 2.33. Experimental coevolution of mutualistic 2 different communities for increased production of riboflavin. (A) Co-culture of a small mutualistic community for increased riboflavin production consisting of *Δrib4:rib5* with either *L. lactis* strain 261 (A) or *B. casei* strain 303 (B). Twelve identical populations were evolved in parallel in 96-well plates. The $\log_2$ fold change of the riboflavin concentration compared to the parental was estimated from cell-free supernatant with fluorescence measurement.

Next, after the end of the serial transfers the final populations of *L. lactis* and *B. casei* that were tested for their riboflavin producing ability by measuring the fluorescence intensity of their conditioned medium on a plate reader. In total, 97 evolved strains of *L. lactis* and 77 strains of *B. casei* were tested, with the percentage of improved strains varying between the species. Approximately 78% of the *L. lactis* isolates exhibited an improved phenotype, with the increase in fluorescence ranging between 0.12 and 2.5 $\log_2$fold change compared to the parental strain. This range is the same as we described for *L. plantarum*, but only 17 strains (17.5%), exhibited an increase equal to or higher than $\log_2$fold 1. In contrast, the same

percentage for *L. plantarum* was 42.7%, a fact that might be connected to the increased number of poor riboflavin producers that arose through the evolution process. The percentage of improved strains in the case of *B. casei* was quite low, as only 44% of the isolates performed better than the parental strain. The small percentage of improved strains can be potentially connected to the fact that this type of community reached high values of OD_{600} fast and thus the selection pressure was comparably low.

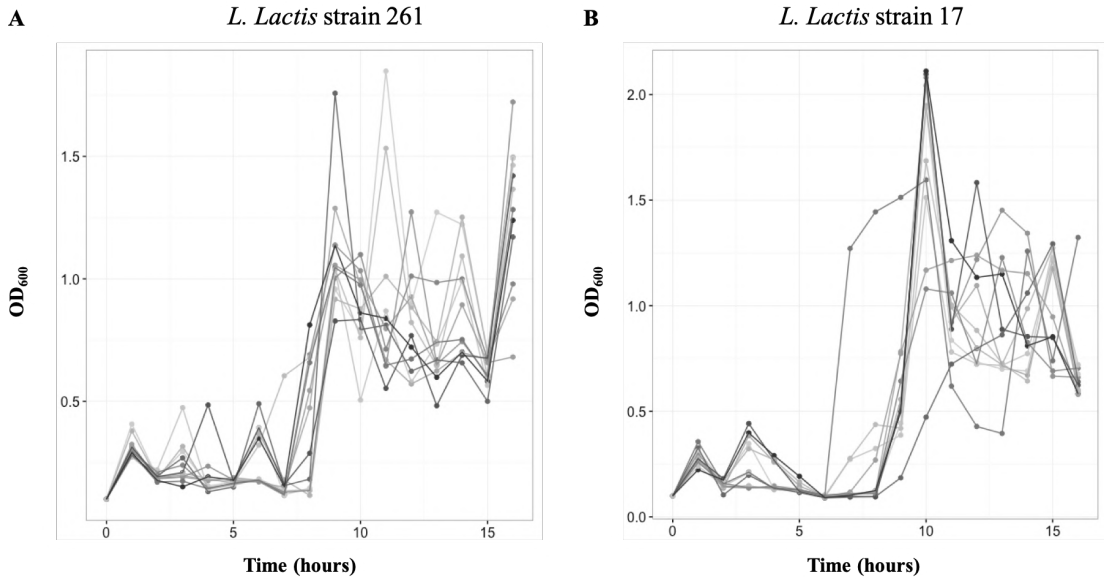

Figure 2.34. Experimental coevolution of mutualistic 2 different communities for increased production of folate. (A) Co-culture of a small mutualistic community for increased folate production consisting of *Δabz1* and either *L. lactis* strain 261 (A) or 17 (B). Twelve identical populations were evolved in parallel in 96-well plates. The growth of each community improved similarly for every replicate, based on optical density measurements.

It seems that there is a trade-off between the number of strains with strongly improved phenotype and the total percentage of improved strains (Figure 2.35). It seems that the rise of specialists favours the simultaneous rise of cheaters, while subtler changes in the production phenotype lead to a more homogeneous population of average secretors. The above mentioned experimental results showcase that this platform of community evolution can potentially be applied to multiple combinations of species for increased production of various compounds, as long as the conditions for mutualism are met.

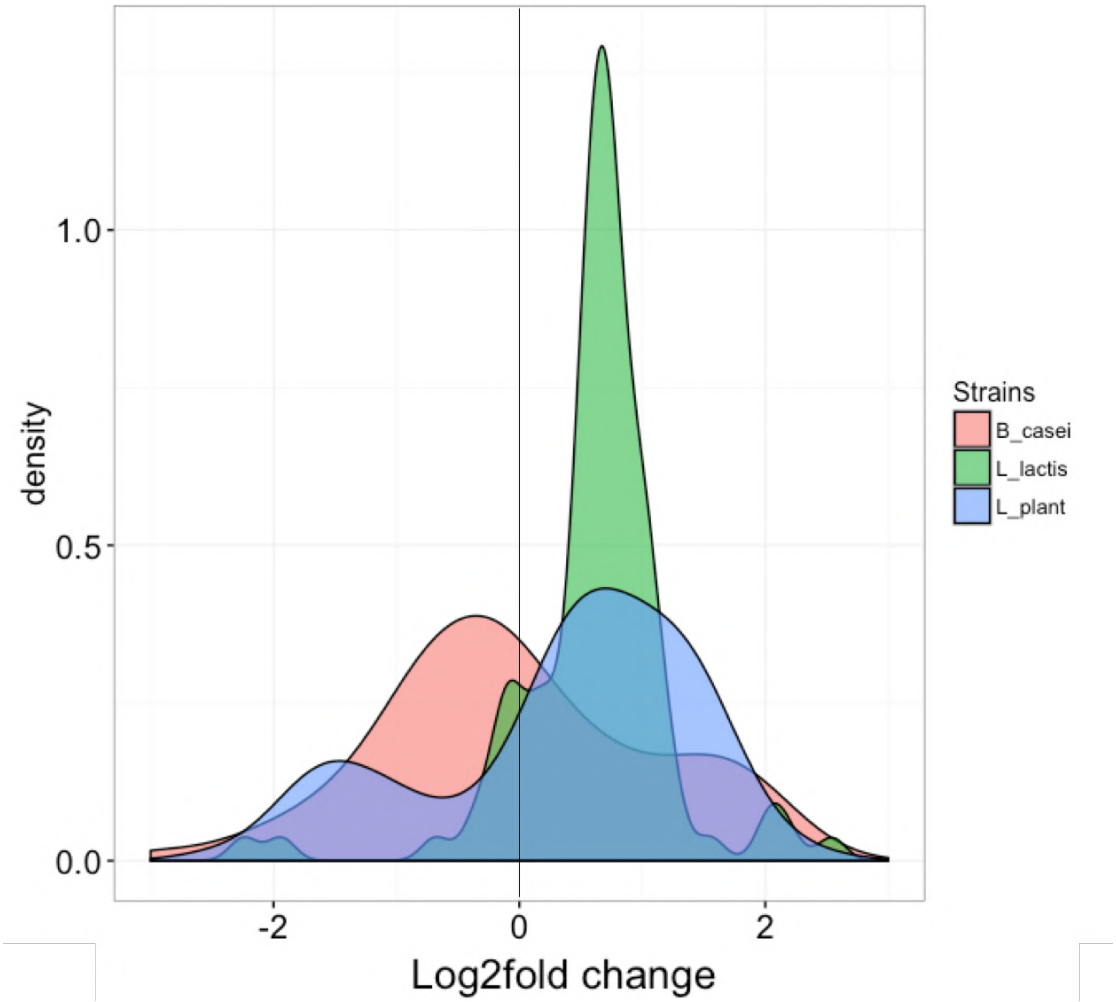

Figure 2.35. Comparison of the efficiency of the coevolution for riboflavin production when different bacterial species are used. Density plot comparing the improvement of evolved isolates originating from similar co-culture experiments for increased riboflavin production. The three species being compared are *B. casei* (77 isolates), *L. lactis* (97 isolates) and *L. plantarum* (134 isolates).

2.2.9 Summary and outlook

Over the past years, the focus of microbiologists has shifted from the study of single species to microbial communities and the interactions between the species that construct these consortia. Even though researchers have made leaps forward in understanding microbial communities there are still puzzling questions that remain unanswered. Such a question is how mutualistic interactions appear in nature. It has been suggested (Harcombe *et al.*, 2018) that mutualism may initially arise through the secretion of metabolic by-products from one species, which are required by a second species for its survival and vice versa, thus creating a dependency between these two species. In the presented study, we took advantage of this type of relationship between species to propose a novel system termed *Syn-A* (from the Greek words synergasia = cooperation and agasti = admirable), which is based on the design of a synthetic microbial community, with the fitness of the community being coupled to the increased production of a target metabolite by one of the species. Coupling the production of the desired compound with the cross-feeding interactions of the community members and subsequently their ability to exhibit improved growth, enabled us to perform ALE to optimise the production and the secretion of the desired compound. Increased concentration of the target compound in the environment will result in improved community fitness, despite of potentially increased production costs for the secreting cells.

We set out to prove our theory by establishing a synthetic microbial community between *S. cerevisiae* and species of LAB, previously studied in our group (Ponomarova *et al.*, 2017), which were evolved towards the increased production and secretion of B group vitamins from non-GMO LAB strains. Increased production of B group vitamins is also important for food biotechnological applications, since the market for functional food, meaning food products with enhanced nutritional value, will have an estimated market share of $275 billion by 2025 (Grand View Research, 2019). Our experiments showed that about two thirds of the strains isolated from mutualistc communities, after evolving together with their yeast partner for at least 80 generations in the case of folate and up to 160 generations in the case of riboflavin, exhibited an improved secreting phenotype. More specifically, strains that were evolved for improved production of riboflavin, secreted up to 6 times more than their parental. The intracellular pools of the vitamin also increased, with the measured amounts ranging from 1.6 to 5.5 times more than the parental strain. Apart from increased amounts of riboflavin, the evolved strains produced increased amounts of FMN and FAD, with both the parental and the evolved strains exhibiting the ability to secrete FAD in the environment. We were able to isolate strains that secret more FAD than riboflavin in

comparison to the parental strain, even though the reason this change occurred remains unclear, since uptake of extracellular FAD has not been proven for *S. cerevisiae*. A potential explanation is that intracellular accumulation of FMN and FAD might lead to a negative feedback loop for riboflavin production and thus FAD secretion may prevent this.

Furthermore, we were able to exhibit that the secreting phenotype arises, despite the potential metabolic burden from increased riboflavin production, by correlating production cost with the evolved strains' decrease in growth rate in monoculture. It seems that through the evolution process the bacterial cells adapt to a strategy that favours the improved growth of the community over the growth of the individual. This was shown in the co-culture growth assays, where communities that include bacteria with a higher secreting phenotype were able to reach a higher maximal OD_{600}, compared to the community that included the parental strain. Moreover, the adaptation towards better community growth was shown with the detection of higher number of RFP positive yeast cells in these co-cultures, as well as with the conditioned medium experiment, where auxotrophic *S. cerevisiae* cells exhibited improved growth in medium from monocultures of evolved strains. It is possible that the increased cost of riboflavin production is balanced by improved growth under conditions of decreased amino acid availability, since in our evolutionary niche the required amino acids were available only after they are secreted from *S. cerevisiae*. Indeed, evolved strains exhibit improved fitness both in yeast conditioned medium and CMD46R- which is was supplemented with eight times less required amino acids than normal.

The observation that the evolved strains exhibit improved fitness in conditions that better simulate the evolution niche (limited availability of amino acids) is important, because it suggests that under these conditions improved riboflavin production is a burden that cells can bear without dire consequences for their overall fitness. Therefore, the question why overproducing phenotypes are favoured and frequently arise through the evolution process can be answered more comprehensively, if we hypothesise that phenotypes which co-operate more, also receive higher rewards from their partners, and as a result, they improve their fitness in the evolution niche. By secreting more riboflavin, lactic acid bacteria increased the abundance of *S. cerevisiae*, consequently increasing the availability of the required amino acids. Similar observations to ours have been previously reported in literature. Zhang and Reed (2014) used two *E. coli* strains with each one being auxotrophic for a single amino acid, provided by the other and after a serial transfer experiment they observed that the evolved isolates had decreased growth rate in monocultures, while the estimated secretion and uptake rate had increased for each auxotroph. Elsewhere, Harcombe *et al.* (2018) created a

mutualism between *E. coli* and *S. enterica* and saw that about 80% of the *S. enterica* exhibited the cooperating phenotype, while the growth rate of the co-cultures improved. Interestingly the authors of this publication discuss how mutualistc evolution can potentially prove useful to generate increased microbial secretion, as it was showcased in our study.

Using different omics approaches we tried to understand the underlying changes that occurred in the bacterial cells during the evolution process and which led to the improved secreting phenotype, as it is summarised in Figure 2.36. Through whole genome analysis we mostly identified mutations that occurred to either regulatory elements such as transcription regulators or transporters, like amino acid transporters. These findings can be associated with what Turkaslan *et al.* (2017) reported after analysing data from the evolution of two obligatory mutualistic species, viz. *Desulfovibrio vulgaris* and *Methanococcus maripaludis*. The researches mostly identified mutations in the coding regions of regulatory genes, while the GO terms analysis they performed was enriched for transcription regulation and two-component sensor activity. Therefore, it seems likely that fine tuning of regulatory elements is important when species adapt to a novel environment and to the available resources, such as optimizing the regulation for using compounds secreted from their partner, while they down-regulate functions that are no longer required.

It seems that higher abundance of the enzymes that participate in the riboflavin biosynthetic pathway is enough to justify the overproducing phenotype. It has been described before (Burgess *et al.*, 2004) that the reactions catalysed by RIBA are the limiting step for riboflavin production in *B. subtilis* and higher abundance of this protein led to a 25% increase in riboflavin yield, but the reported increase was not as significant in *L. lactis*, which needed all the biosynthetic genes to be overexpressed. It seems that the *L. plantarum* strain that we used is similar to both of these cases, considering that we detected higher abundance of RIBA in all the evolved strains, but for each one of them we observed higher amounts of two or three more biosynthetic genes. Nevertheless, in our study we describe an increase of extracellular riboflavin equal to 200%, which cannot be justified solely by the increase in biosynthetic enzymes. One plausible explanation is that during the evolution process the bacterial cells had the opportunity to upregulate other pathways that are associated with the biosynthesis of riboflavin, to increase the availability of the required precursors. For example, a different engineering strategy that was adopted for riboflavin overproduction in *B. subtilis*, was to divert fluxes through the purine biosynthesis, by overexpressing five genes of the *pur* operon (*purF*, *purM*, *purN*, *purH*, and *purD*), aiming for increased production of GTP through reactions that use glutamine or glycine as co-substrates (Schwechheimer *et al.*,

2016). Interestingly the proteins encoded by these genes were upregulated in the tested evolved strains, while metabolomics analysis showed that these two amino acids were abundant inside the cells. Moreover, the protein PURR, which is the repressor of the *pur* operon (Schwechheimer *et al.*, 2016) was downregulated in all the evolved strains, while the detected higher levels of ribose-5phosphate hint that in the evolved strains the glucose metabolism was diverted towards the PPP. This potential diversion of the metabolism can also be linked with higher availability of precursors, as well. In conclusion, our data suggest that the optimization of only one pathway is not sufficient for the expression of an overproducing phenotype, but changes in different aspects of the metabolism seem to act synergistically. Finally, it is worth to mention that through different approaches (proteomics and metabolism) the evolved strains tested formed the same clusters. These clusters were in agreement to the riboflavin producing phenotype of the strains with the two main secretors clustering together, while the other two strains which secrete mainly FAD formed the second cluster.

One major advantage of our proposed system *Syn-A* is that it is directed and specific. Our data suggest that the "first rule" of ALE which states that "you get what you select for" holds true for our system, since the evolved strains selectively overproduce the compound for which we created a selection pressure. The higher amounts of produced FAD should not be considered as divergence from the desired outcome, since it can potentially be taken up by the yeast, or again by the bacteria and be recycled to produce riboflavin. Furthermore, testing the riboflavin secreting ability of strains evolved to overproduce folate, showed that the conditions we create lead to specific results, while NGS showed that evolution under obligatory mutualism leads to similar adaptations, such as mutations in regulatory genes. The serial transfer experiment of *L. plantarum* monocultures, exhibited that it is in fact mutualism that creates the pressure for increased production and not the lack of a specific vitamin from the environment. Finally, we showed that our proposed method is not limited by the selection of species, as long as the conditions for metabolic cross feeding are met, meaning that the use of our system can potentially expand to the production of other compounds, apart from vitamins.

Bacterial strains that produce vitamins have been successfully used for a variety of applications. First of all, both engineered strains and strains that were exposed to roseflavin have been used to produce food with increased levels of riboflavin, such as dairy products, bread and pasta. Similarly, bacterial strains that produce folate have been successfully used to increase the concentration of this vitamin in yoghurt and cereals (LeBlanc *et al.*, 2020).

Other studies have shown that when either the bacterial strains or the fortified food products were fed to mice with vitamin deficiencies the phenotype returned to normal and the manifestation of ariboflavinosis (lack of riboflavin) was eliminated (LeBlanc *et al.*, 2005, Carrizo *et al.*, 2020). Moreover, vitamin producing LAB strains have been used to treat animal models of inflammatory bowel disorders (IBD) (LeBlanc *et al.*, 2020). For example, a riboflavin producing strain of *L. plantarum* was used to decrease the levels of inflammation in mice with colitis (Levit *et al.*, 2017). Targeted improvement of the vitamin producing abilities of LAB with non-GMO technologies, like the newly proposed *Syn-A* can prove important, considering that vitamin deficiencies remain an actual problem in both developing and developed societies (LeBlanc *et al.*, 2005) and that there are globally almost seven million cases of IBD (LeBlanc *et al.*, 2020).

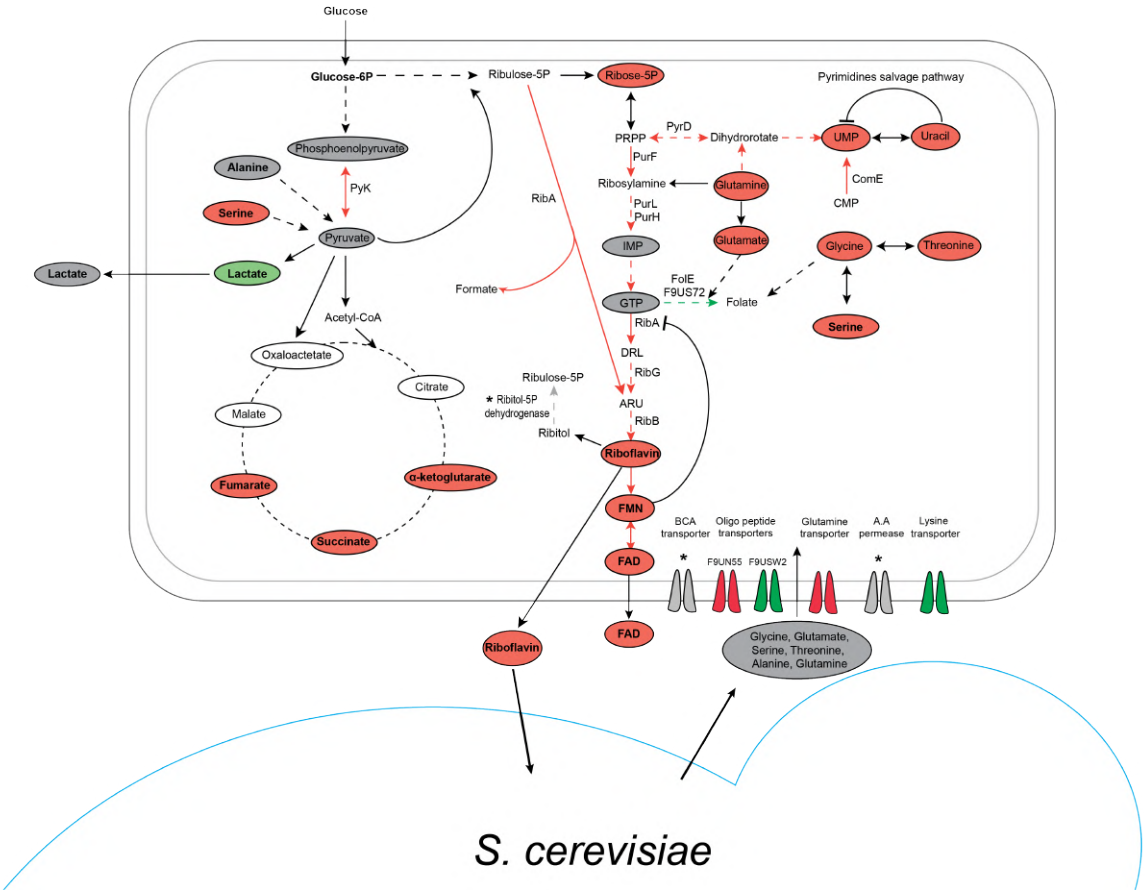

Figure 2.36. Schematic representation of the metabolic changes that occurred to the bacterial cells during the mutualistic evolution. Eclipses surrounding the name of a compound are used for evidence collected through metabolomics analysis. Proteomics evidence are represented as differently colored arrows, with the name of the respective protein close to the arrow. Evidence from NGS are represented as an asterisk. Solid arrows inside the cell represent reactions that occur through one enzymatic step, while dashed arrows represent several intermediate reactions. Upregulation of proteins and increased amounts of metabolites are represented with red color, downregulation and decreased amounts of metabolites are represented with green color, while non-detected differences are represented with grey color.

2.3 Material and methods

2.3.1 Strains and culture conditions

In the present study kefir isolates of lactic acid bacteria (*Brevibacterium casei*, *Lactococcus lactis* and *Lactobacillus plantarum*) were used (Table 2.5). Prior to the start of experiments each of the species was pre-cultured overnight in appropriate rich medium which was sterilized through autoclaving. *L. plantarum* was grown in MRS, *L. lactis* in GM17 and *B. casei* in MGAM. These media were prepared according to the manufacturer's instructions. All the bacteria cultures were grown statically at 30 ^{o}C. All three-bacterial species were provided by Dr. S. Blasche (Patil group, EMBL Germany). Prototrophic strains of *Saccharomyces cerevisiae* strain S90 were grown overnight in rich medium (YPD) containing 10 g/L of yeast extract, 20 g/L peptone and 20 g/L glucose which was sterilized by autoclaving. YPD was further supplemented with 1 µg/mL of riboflavin when it was used for growth of auxotrophic yeast cells.

Table 2.5. List of species used in the present study.

Species	Source
Lactobacillus plantarum strain GKC	K.R Patil Laboratory
Lactococcus lactis strain 261	K.R Patil Laboratory
Lactococcus lactis strain 17	K.R Patil Laboratory
Brevibacterium casei strain 303	K.R Patil Laboratory
Saccharomyces cerevisiae strain S90	Steinmetz *et al.*, 2002

The chemically defined medium (CDM35) was described by Ponomarova *et al* (2017) and was designed by reducing the number of components in the rich medium composed as a hybridn of previously described media. Two different versions of the medium were used, in order to create the proper evolution niches. The first one is CDM34R-, for which riboflavin was omitted and the other one is CDM33F -, from which both folate and paraminobenzoic acid were omitted.

Single bacteria cultures are supplemented with 40 mL/L of amino acids solution, defined as CDM46R-, in cases riboflavin was omitted or CDM45F - when folate and para-aminobenzoic acid were omitted. The composition of this solution is shown in table 2.6.

Table 2.6. List of the supplemented amino acids. The name and the concentration of amino acids, that need to be supplemented in CDM35 to support the growth of LAB in monoculture.

Amino acid	Concentration (g/L)
L-Glutamic acid	15
L-Phenylalanine	10
L-Proline	17.5
L-Asparagine	12.5
L-Aspartic acid	10.5
L-Glutamine	15
L-Serine	12.5
L-Threonine	12.5
L-Cysteine	5
L-Alanine	10
Glycine	7.5
L-Lysine	10
L-Tryptophan	5

Liquid medium experiments with single bacteria strains or yeast mutant strains were started at 0.01 OD_{600} using pre-cultures washed 3 times with phosphate buffered saline (PBS). The washed yeast cells were starved for 3 h at RT prior to the start of the experiment. All cultures were grown statically at 30 °C. To measure the growth rate and vitamin production of single strains, they were grown statically in 50 mL Erlenmeyer flasks, containing CDM46R- or CDM45F-, unless a different recipe of CDM is specified. OD_{600} measurements were done using a spectrophotometer.

2.3.2 Yeast strain engineering and auxotrophy construction

Riboflavin auxotrophy in the yeast strain S90 was introduced with two single deletions. Firstly, gene *RIB4* which encodes for the protein Lumazine synthase was deleted, as lumazine is an immediate precursor of riboflavin. Positive clones were selected and were sequentially subjected to a second round of engineering, with the deletion of gene *RIB5*, which encodes for riboflavin synthase, an enzyme that catalyses the last step of the riboflavin biosynthetic pathway. For the evolution of folate overproduction, we deleted *ABZ1*, gene that encodes for para-aminobenzoic synthase. The selected genes were amplified from the corresponding strains of the prototrophic *S. cerevisiae* gene deletion collection (Mülleder *et al.*, 2012) and were inserted to 2 different plasmids, puG6 (Güldener *et al.*, 1996) and puG32 (Goldstein *et al.*, 1999) (Table 2.7). Fragments containing either a kanMX4 or a hphMX4 cassette were amplified by PCR and the products were used to transform wild-type *S. cerevisiae* S90. For the transformation, yeast cultures in mid exponential phase (OD_{600} = 0.7,

50 mL) were centrifuged, washed, and resuspended in 1 mL of sterile water. Then 100 µL of the cell suspension were combined with an appropriate volume of transformation mix, which consisted of 240 µL PEG 3500 (50% w/v), 36 µL lithium acetate (1.0 M), 50 µl boiled single- stranded carrier DNA (2 mg/mL), 20 µL of PCR amplification product and 74 µL of ddH_2O. The cells with the transformation mixture were heat-schoked for 40 min at 42 ^oC in a water bath and then they were re-suspended in YPAD and incubated for 3-4 hours to allow for the expression of the integrated anti- biotic marker. Finally, the cells were platted on YPAD agar plates containing the appropriate antibiotic (G418 300 µg/mL, Hygromycin 150 µg/mL) supplemented with either riboflavin or folate and single colonies were selected. The success of homologous recombination was verified by colony PCR.

For the constitutive expression of RFP in the riboflavin auxotrophic yeast we used the vector pCfB2049, described in Stovicek *et al.* (2015). The same transformation protocol was used, but in this case the antibiotic of choice was nourseothricin sulfate (clonNat, 100 µg/mL). All the plasmids and the primers used in this study are described in Table 2.7.

Table 2.7. List of plasmids used in the present study.

Plasmid	Gene	Primers (5'-3')
puG6	*RIB4*	Forward: GCAGTATAACGCAGTATAACGCAGTATAACGCAGTAGCGACATGGAGGCCCAGAA Reverse: GCGCTTATTCAAAAAGCATTTTTACCGAACTTAACTCGACACTGGATGGCGGCGT
puG32	*RIB5*	Forward: ATGTTTACTGGTATTGTAGAATGCATGGGGACTGTAGCGACATGGAGGCCCAGAA Reverse: GGTAGTTTCTAACCTTCTCCTCGATAATGTTTGAGTCGACACTGGATGGCGGCGT
puG6	*ABZ1*	Forward: ATCACCTACACTATCTTCAGCAAGG Reverse: CTTTAGGTACGGTTGTTGTCATCTT
pCfB2049	NA	Forward: TGACGAATCGTTAGGCACAG Reverse: CCGTGCAATACCAAAATCG

2.3.3 Adaptive laboratory evolution

In the present study, six similar mutualistic communities were constructed, each containing either different species of LAB or different engineered yeast strains. The first community consisted of Δ*rib4:rib5 S. cerevisiae* and *L. plantarum* for increased production

of riboflavin, the second one consisted of Δ*abz1 S. cerevisiae* and *L. plantarum* for increased production of folate. The rest contained Δ*rib4:rib5 S. cerevisiae* and either *L. lactis* or *B. casei* for production of riboflavin or Δ*abz1* and two different strains of *L. lactis* for the production of folate. At the beginning of the evolution experiment single bacteria colonies were pregrown as described previously and were mixed with the appropriate auxotrophic yeast strain. In total 12 co-cultures of each community were started in parallel, with each containing both species in 1:1 volume ratio based on their OD_{600}. The initial OD_{600} value of the communities was equal to 0.01 and they were inoculated in 96-well plates containing 200 μL CDM34R- or CDM33F-. The evolution was performed under microaerobic conditions (statically) at 30 ^{o}C. Growth was monitored in a Tecan microplate reader and when the communities reached stationary phase, the appropriate volume from each well was transferred to fresh medium with a starting OD_{600} value of 0.01. Stability of the communities was estimated by platting on YPAD plates, where all members could grow and form colonies. Intermediate and final populations were stored by adding 80% [w/v] glycerol directly to the plate wells (final concentration 30% [w/v]) and then storing it in -80 ^{o}C. The total duration of the experiment and the number of generations varied between species and between communities. *L. plantarum* communities were evolved for riboflavin production for 3 months and during this period the populations were transferred 24 times, while a total number of 160 generations passed. Estimation of the number of generations that passed is based on the measured OD_{600}. The evolution of *L. plantarum* for folate production lasted 2 months, with a total of 15 transfers and 80 generations. Evolution of *B. casei* and *L. lactis* lasted for a total of 80 days and approximately 140 generations in the case of riboflavin, and 80 generations in the case of folate, passed.

For the monoculture ALE experiment of *L. plantarum* in CDM46R-, we followed the same procedure with the difference that we started with 4 replicates in parallel, instead of 12 like in all the previously described community evolutions. This evolution lasted for 2 months and in total 100 generations passed.

2.3.4 Estimation of riboflavin secretion through fluorescence intensity assay

All final populations from each community were platted on solid media, each composed of the appropriate selective medium and 2% [w/v] agar as solidifying agent. From each population, seven single colonies were picked and transferred to a separate well of a 96-well plate filled with 200 μL of CDM46R-. In a similar fashion 4 singles colonies of the corresponding parental strain were isolated and inoculated in the same 96-well plate, while a well of CDM46R- was used as an empty control. The plate was incubated for 72 hours, which

was on average the time point between the passages during the serial transfer experiment. After the 72 hours passed the plate was centrifuged for 3 min at 1500 rpm in an Eppendorf centrifuge and the top 100 µL of supernatant were transferred to a black opaque 96-well plate for fluorescence measurement. The measurements were done in a Tecan plate reader, using the 440 nm excitation at medium intensity and detecting at 544 nm. The average fluorescence intensity for the parental strain was calculated and this value was used to calculate the log_2fold change of the evolved strains.

2.3.5 Ultra-Performance Liquid Chromatography data acquisition and analysis

Riboflavin was extracted from 1 mL of culture supernatant by addition of an equal amount of 20% acetonitrile: H2O and addition of 1 µg/mL of roseflavin as internal standard. The samples were left at 4 ^{o}C for 10 min and then were transferred to dark glass vials. Next, the samples were concentrated using a vacuum concentrator, at 30 ^{o}C for 2 hours and subsequently reconstituted in 150 µL of 20% acetonitrile:H_2O. The same process was followed for the identification of folate content in extracellular samples. In the case of measuring intracellular samples, 1 mL of cell culture was pelleted through centrifugation (2 min at 15000 rpm) and the pellet was washed once with PBS. The cell pellets were lysed by sonication (3 rounds of 8 seconds, maximum intensity) and a heating step (72 ^{o}C for 15 minutes). The reconstituted samples were filtered through a 0.22 µm syringe filter.

The liquid chromatography method for the detection of vitamins was run on a Waters Acquity UPLC H-Class instrument with a PDA detector and a quaternary solvent system. The established method was 8 minutes long, with a flow rate of 0.35 mL/min and run on a BEH Hilic column (130 Å, 1.7 µm, 2.1 mm X 150 mm, Waters, part number 176001094). The column was heated to 35 ^{o}C and samples were kept at 6 ^{o}C and 5 µL of sample was injected. The method used 50% [v/v] acetonitrile (CAN; Biosolve, ULC grade) as a washing buffer, and 50% [v/v] methanol (Biosolve, ULC grade) as purging buffer. The organic mobile phase was acetonitrile, 10 mM ammonium acetate (Sigma-Aldrich, HPLC grade) was used as buffer and the pH was adjusted to 9. The two buffers used were 50% ACN + 10mM Ammonium acetate pH 9 and 100% ACN + 10mM Ammonium acetate pH 9. The full method is described on Table 2.8 and the elution time of every detected compound on Table 2.9. The elution time of each vitamer was estimated by using pure grade standards of riboflavin (Sigma-Aldrich), FAD (≥95%, Sigma-Aldrich), FMN (≥95%, Sigma-Aldrich), folate (≥97%, Sigma-Aldrich) dihydrofolate (≥90%, Sigma-Aldrich) and tetrahydrofolate ((≥65%, Sigma-Aldrich) and is presented in Table 2.9.

Table 2.8. Description of the UPLC method used for detection of vitamins.

Time (minutes)	Flow (mL/min)	Acetonitrile 100% + 10mM Ammonium acetate, pH 9	Acetonitrile 50% + 10mM Ammonium acetate, pH 9
0-3	0.35	10	90
3-5	0.35	70	30
5-7	0.35	10	90
7-8	0.35	19	90

Table 2.9. List of the compounds detected with UPLC and their retention time.

Compound	Peak elution time (min)
Roseoflavin	3.5
Riboflavin	3.9
FAD	5.9
FMN	6.4
Folate_A	4.36
Folate_B	4.6
DHF_A	3.95
DHF_B	4.35
DHF_C	4.6
THF_A	3.95
THF_B	4.35
THF_C	4.6

All chromatograms were annotated with the vendor specific program Empower 3, and manually curated for peak identification. The readout for all chromatograms is the baseline corrected area under the curve of the vitamin peak and the internal standardpeak respectively. In the case of folate every peak of the standards was taken separately into consideration separately.

2.3.6 Liquid Chromatography –Mass Spectrometry data acquisition and analysis †

2.3.6.1 Liquid Chromatography –Mass Spectrometry data analysis

The metabolite extracts were separated using HILIC on a 2.1 x 100 mm, 3.5 µm BEH amide column (Waters). Separation was achieved by gradient elution (100% B to 100% A in

† This method part was provided by Rose Gathungu, Metabolomics Core Facility, EMBL, Heidelberg, Germany.

20 min) on a binary solvent Exion LC system (Sciex, Darmstadt, Germany). Mobile phase A consisted of 80:20 H_2O:ACN and mobile phase B consisted of 85:15 CAN: H_2O. Both mobile phase A and B were modified with 10 mM ammonium acetate and adjusted to pH 8.5 with 25% [v/v] ammonium hydroxide. A flow rate of 300 µl/ minute was used for separation and the column and sample tray were held at 25 °C and 4 °C respectively. A 5 µL injection volume was used for all samples.

2.3.6.2 MS analysis

MS analysis was performed in the MRM-scan mode on a Q-Trap 6500+ (Sciex, Darmstadt, Germany) equipped with an Ion Drive Turbo V electrospray ionisation probe. In both the positive and negative ionisation mode, the source temperature was set to 550 °C, the curtain gas was set to 30, the nebulizer gas (GS1) and the auxiliary gas (GS2) were both set to 45, and, the collision activation dissociation (CAD) gas was set to medium. Nitrogen was used as both the curtain gas and the CAD gas. In the positive ionization mode, the spray voltage was set to 5.5 kV, the declustering potential to 90, and the entrance and exit potentials to 10. In the negative ionization mode, the spray voltage was set to -4.5 kV, the declustering potential to -60 and the entrance and exit potentials were both set to -10. Both the LC and MS were controlled using Analyst software version 1.7.

The MRM transitions used for analysis of the vitamins were: Folate 442/295.1, Riboflavin 377/243 and FAD 786/348 all analysed in the positive ionization mode, and, for FMN 455/213 analysed in the negative mode. The MRM transitions used for the analysis of the nucleotides were as follows: IMP 349/137, UMP 325/97, GMP 364/154, AMP 348/136 all in the positive ionization mode, and, Uracil 111/42 analysed in the negative mode.

Data analysis was performed using the Sciex OS software. A Gaussian smoothing of 1 point was applied to all peaks.

LCMS experiments were performed at the Metabolomics Core Facility (EMBL Heidelberg).

2.3.7 Growth kinetics in CDM46R- and CDM46F-

Biological triplicates of the parental and the evolved strains were inoculated in rich medium and were left to grow overnight. These overnight cultures were washed three times with 1X PBS and were starved in PBS for 2 hours. Then, cells were inoculated with an initial OD_{600} of 0.01 in a final volume of 7 mL in 50 mL Erlenmeyer flasks and were grown statically in a 30 ^{o}C incubator. Estimation of growth was based on increase of the optical density which was measured at regular time intervals. Growth was followed until the OD_{600}

remained stable, the time point where cells had reached stationary phase. Growth rates were estimated by calculating the maximum slope values of the best fit to log2 transformed OD values. Between 4 and 5 points were used for these calculations. Estimation of the growth curve in CDM46R- with diluted amino acids was performed in a similar fashion. In this experiment, three biological replicates of the parental and the selected evolved strains were inoculated in 200 µL of CDM46R- in a 96-well plate. CDM46R- was supplemented both with 40 mL/L and with 5 mL/L of amino acid solution (Table 2.6), in technical replicates. The 96 well plate was sealed with a Breathe-Easy® sealing membrane (Z380059, Sigma-Aldrich) and growth was followed on a Tecan microplate reader at regular time intervals.

2.3.8 Conditioned media assays:

2.3.8.1 Bacterial conditioned medium

Bacterial strains were cultured in CDM46R- medium until mid-exponential phase, the cultures were centrifuged for 3 min. at full speed)and then the supernatant was filtered through a 0.22 µm syringe filter. Auxotrophic yeast cells were from an overnight culture were washed 3 times in 1X PBS, were starved for 3 hours and then inoculated in the conditioned medium without any further supplementation. The initial OD_{600} value was 0.01 in a final volume of 200 µL in a 96 well plate, which was sealed with a Breathe-Easy® sealing membrane. Growth was monitored in a Tecan microplate reader, with OD_{600} measurements every hour for a total of 48 hours. After the cultures had reached stationary phase the 96 well plate was thoroughly mixed and the OD_{600} was measured one last time. Growth rates were estimated by calculating the maximum slope values of the best fit to log2 transformed OD values. Between 6 and 8 points were used for these calculations.

2.3.8.2 Yeast conditioned medium

Wild type *S. cerevisiae* S90 cells were cultured in CDM34R-, until they reached mid-exponential phase. Then, the cultures were centrifuged at full speed for 3 min and the medium supernatant was filtered with 0.22 µm syringe filters. Bacterial cells (parental and evolved) from overnight cultures were washed 3 times in 1X PBS and were starved for 2 hours. Then they were inoculated with an initial OD_{600} of 0.01 in a final volume of 200 µL in a 96-well plate, which was sealed with a Breathe-Easy® sealing membrane. All bacterial strains were inoculated in conditioned medium originating from the same yeast culture.

2.3.9 Flow cytometry

Total volume equal to 0.5 mL of *L. plantarum* and *S. cerevisiae* co-culture was pelleted. The pellet was fixed with 4% [w/v] formaldehyde-PBS solution at RT for 10 min. The fixed pellet was washed once with PBS and then rehydrated in phosphate buffered saline with 1 mM EDTA. Prior to flow cytometry the pellets were briefly sonicated 3 times (10 seconds with 0.5 second ON-OFF intervals; 10% amplitude; Branson Sonifier W-250 D, Heinemann) at 4 ^{o}C. Samples were analyzed by flow cytometry using a Cytek Aurora Cytometer (Cytek) at low flow. A total volume of 35 µL was analysed for each sample, thresholds used were 800 for forward scatter height and 600 for side scatter height. Yeast cells were identified based on the RFP intensity and cells under the intensity threshold of 10^4 were excluded from the analysis. Flow cytometry was performed at the Flow Cytometry Core Facility (EMBL Heidelberg).

2.3.10 Microscopy

Brightfield and fluorescence images were acquired at the Advanced Light Microscopy Facility at EMBL, Heidelberg with a ZEISS Cell Observer Widefield microscope, equipped with a metal halide fluorescence lamp. Observations were made using a 63× oil objective and the acquired images were analysed with ImageJ (Schindelin et al., 2012).

2.3.11 Genomic DNA extraction from Lactic acid bacteria

Overnight bacterial cultures of 4 mL total volume were pelleted in 2 mL Eppendorf tubes and were resuspended in 1 mL of TES buffer (Tris50 mM, EDTA 20 mM and sucrose 2 mM) which contained 20 mg/mL of lysozyme. Then, the tubes were incubated for 30 min at 37 ^{o}C. The cell suspensions were transferred to FastPrep Cap tubes with 200 µL of glass beads (150-212 nm acid washed, Sigma) and the cells were mechanically disrupted with 3 rounds of bead beating at 4.5 Mhz/sec for 20 seconds with 1 minute cooling intervals. Then 150 µL of 20% [w/v] SDS was added to the tubes and they were incubated for 5 min at RT, before being centrifuged for 1 min at full speed. The supernatant was collected in a fresh tube and 100 µL of proteinase K (20 mg/mL) was added and the samples were incubated for 30 min at 37 ^{o}C, followed by the addition of 200 µL of 5 M pottasium acetate and incubation for 15 min on ice. The tubes were centrifuged for 15 min at max speed at 4 ^{o}C and the supernatant was transferred to a new tube in a 1:1 ratio with phenol-chloroform/isoamylic alchocol. Then, the tubes were centrifuged briefly until an emulsion was formed. The emulsions were centrifuged at 13.000 rpm for 10 min at RT and the aqueous phase was transferred to a new

tube, mixed with an equal volume of ice cold 100% ethanol and DNA was precipitated overnight at -20 °C. Next day the tubes were centrifuged for 10 min at max speed at 4 °C and the DNA pellets were washed with 70% [v/v] ethanol and then dried for 20 min at 30 °C. The dried pellets were dissolved in 60 µL of ddH_2O with 100 µg/mL RNAse and were incubated at 37 °C for 15 min. The quality of the extracted DNA was by agarose gel electrophoresis (1% [w/v]). DNA concentrations were measured using Qubit (Thermo Fisher Scientific, USA). Dissolved DNA samples were stored at 4 °C until further processing.

2.3.12 Whole genome sequencing and data analysis

Equal amounts of DNA from all samples were used for library preparation, which was done with the NEBNext DNA Ultra2 Library Preparation Kit (New England Biolabs). The preparation of the library was performed on an automated liquid handling system (Hamilton Robotics), the quality of the library was tested on a 2100 BioAnalyzer (Agilent Technologies), and the DNA concentration was measured using a Qubit. Sequencing was performed at the Genomics Core Facility (EMBL Heidelberg) with use of the HiSeq2500 platform (Illumina, San Diego, USA) and the run produced 250 bp paired-end reads.

Identification of mutations that occurred during the ALE co-culture experiment was performed with Breseq (Deatherage and Barrick, 2014). Breseq is a computational pipeline for the analysis of short read re-sequencing data and it uses reference-based alignment approaches for its predictions. Therefore, we decided to compare the sequence of the parental strain with evolved isolates, but the parental *L. plantarum* strain used in this study had not been previously sequenced. As a result, the first step before the use of Breseq was to assemble the parental sequence. The initial quality of the raw Illumina paired-end reads was evaluated with FastQC (Babraham Bioinformatics) and were subsequently trimmed with CUTADAPT (Martin, 2011) setting a quality threshold of 15 and a length cutoff of 150 bp. Next, the pair ended files were merged using PEAR (Zhang *et al.*, 2014) with the default parameters. Assembly of the connected reads was done with SPAdes (Nurk *et al.*, 2013) using the parameters `--careful, -k 5, 15, 35, 55` and all others as default, while the evaluation of the assembly was done with QUAST (Gurevich *et al.*, 2013). The assembled genome was annotated with RAST (Overbeek *et al.*, 2014) through which we identified as the closest related *L. plantarum* strain to ours the strain LP3, for which the whole genome sequence was available in NCBI (direct submission). We used this strain as a reference sequence in Breseq, as its inputs are GFF3 files.

We used Breseq to compare the genome of strain LP3 and the genome of the parental strain, with the default parameters and additionally the commands `–l 120, -j 4.` As it

was expected we were able to identify differences between our parental and strain LP3, which we then applied to the reference genome, with the gdtools' command APPLY, which is provided by Breseq as well. This way we created a "mutated" reference genome which was used for all further comparisons with the evolved strains.

The coverage of reads on the chromosome was performed with Bedtools, with use of the commands "makewindows" and "coverage". Before using bedtools it was necessary to use Bowtie2 to create an alignment of the reads and then Samtools to sort the alignment to coordinates. The sorted bam file was converted into a bed file, before use of the command "coverage". The resulting file was visualised in R.

2.3.13 Proteome analysis

2.3.13.1 Protein extraction

For the extraction of total proteome 10 mL of each culture were transferred into ice-cold 15 mL Falcon® tubes which were centrifuged immediately at 3000 rpm for 3 min at 4^{o}C (Eppendorf centrifuge). The supernatant from the centrifugation was discarded and the cell pellets were washed once with 1 mL of cold PBS buffer. The washed pellets were snapped frozen with liquid nitrogen and stored at -80^{o}C. For the extraction, the cell pellets were lysed with 0.1% RapiGest (Waters) in 100 mM ammonium bicarbonate and mechanical disruption with 3 rounds of sonication of 10 sec sonication and 10 sec rest on ice per round. Sonication was followed by 2 cycles of bead beating (200 µL glass beads, 150-212 nm acid washed, Sigma), each cycle lasting 20 sec at 4 Mz/sec with 1 min cooling intervals between the cycles. Sample preparation of the extracted proteins and initial analysis of the MS data were performed at the Proteomics Core Facility (EMBL Heidelberg).

2.3.13.2 Sample preparation and TMT labelling ‡

Reduction of disulphide bridges in cysteine containing proteins was performed with dithiothreitol (56°C, 30 min, 10 mM in 50 mM HEPES, pH 8.5). Reduced cysteines were alkylated with 2-chloroacetamide (room temperature, in the dark, 30 min, 20 mM in 50 mM HEPES, pH 8.5). Samples were prepared using the SP3 protocol (Hughes *et al.*, 2018) and trypsin (sequencing grade, Promega) was added in an enzyme to protein ratio 1:50 for overnight digestion at 37°C. Peptides were labelled TMT6plex (Dayon *et al.*, 2008) Isobaric Label Reagent (ThermoFisher) according the manufacturer's instructions. For further sample clean up an OASIS® HLB µElution Plate (Waters) was used. Offline high pH reverse phase fractionation was carried out on an Agilent 1200 Infinity high-performance liquid

chromatography system, equipped with a Gemini C18 column (3 µm, 110 Å, 100 x 1.0 mm, Phenomenex), (Reichel *et al.*, 2016), resulting in 12 fractions.

2.3.13.3 Mass spectrometry data acquisition ‡

An UltiMate 3000 RSLC nano LC system (Dionex) fitted with a trapping cartridge (µ-Precolumn C18 PepMap 100, 5µm, 300 µm i.d. x 5 mm, 100 Å) and an analytical column (nanoEase™ M/Z HSS T3 column 75 µm x 250 mm C18, 1.8 µm, 100 Å, Waters). Trapping was carried out with a constant flow of trapping solution (0.05% trifluoroacetic acid in water) at 30 µL/min onto the trapping column for 6 minutes. Subsequently, peptides were eluted via the analytical column running solvent A (0.1% [v/v] formic acid in water) with a constant flow of 0.3 µL/min, with increasing percentage of solvent B (0.1% [v/v] formic acid in acetonitrile) from 2% to 4% in 4 min, from 4% to 8% in 2 min, then 8% to 28% for a further 37 min, in another 9 min. from 28%-40%, and finally 40%-80% for 3 min followed by re-equilibration back to 2% B in 5 min. The outlet of the analytical column was coupled directly to an Orbitrap QExactive™ plus Mass Spectrometer (Thermo) using the Nanospray Flex™ ion source in positive ion mode.

2.3.13.4 MS method for TMT6 ‡

The peptides were introduced into the QExactive plus via a Pico-Tip Emitter 360 µm OD x 20 µm ID; 10 µm tip (New Objective) and an applied spray voltage of 2.2 kV. The capillary temperature was set at 275°C. Full mass scan was acquired with mass range 375-1200 m/z in profile mode with resolution of 70000. The filling time was set at maximum of 100 ms with a limitation of $3x10^6$ ions. Data dependent acquisition (DDA) was performed with the resolution of the Orbitrap set to 17500, with a fill time of 50 ms and a limitation of $2x10^5$ ions. A normalized collision energy of 32 was applied. Dynamic exclusion time of 20 s was used. The peptide match algorithm was set to ‘preferred’ and charge exclusion ‘unassigned’, charge states 1, 5 - 8 were excluded. MS^2 data was acquired in profile mode.[2]

2.3.13.4 MS data analysis ‡

IsobarQuant (Franken *et al.*, 2015) and Mascot (v2.2.07) were chosen for data processing. A Uniprot *L. plantarum* proteome database (UP000000432) containing common contaminants and reversed sequences was used, with the addition of the sequences of the absent riboflavin biosynthetic proteins (ribD and ribF). The search parameters were the following: Carbamidomethyl (C) and TMT10 (K) (fixed modification), Acetyl (N-term),

‡ This method part was provided by Mandy Rettel and Frank Stein, MProteomics Core Facility, EMBL, Heidelberg, Germany.

Oxidation (M) and TMT10 (N-term) (variable modifications). A mass error tolerance of 10 ppm was set for the full scan (MS1) and for MS/MS (MS2) spectra of 0.02 Da. Trypsin was selected as protease with an allowance of maximum two missed cleavages. A minimum peptide length of seven amino acids and at least two unique peptides were required for a protein identification. The false discovery rate on peptide and protein level was set to 0.01.

2.3.14 Gene ontology enrichment analysis

Gene ontology analysis was performed with use of the online version of DAVID 6.8 (Huang et *al.*, 2009). The parameters of the analysis were: significance threshold, Benjamini-Hochberg, p-value <0.05.

2.3.15 Cell surface traits

2.3.15.1 Autoaggregation and coaggregation

Evaluation of the cels aggregation phenotype was performed as described by Kotzamanidis *et al.* (2010). In short, overnight cultures of bacteria and WT S90 yeast in biological triplicates were washed 3 times in 1x PBS. The OD_{600} was measured and an appropriate volume of the washed cultures was transferred to a cuvette to have a starting OD_{600} of 1. For the coaggregation experiment, bacteria and yeast were mixed with a 4:1 ratio (80% bacteria cells and 20% yeast cells) to better represent the higher number of bacteria cells that were detected during co-culture experiments. All the cuvettes from both experiments were left still in a place with no vibrations for the duration of the experiment and the OD_{600} of each one was measured every hour. A final measurement was performed after 24 hours. The percentage of aggregation is calculated based on the decrease of the measured OD_{600} values, using the formula $(1- A_t/A_0) \times 100$, where A_t is the OD_{600} value of every time point and A_0 is the starting OD_{600} value.

2.3.15.2 Microbial Adhesion to Hydrocarbons (MATH)

Estimation of microbial cell surface hydrophobicity was performed with the microbial adhesion to hydrocarbons (MATH) test as describe previously (Kotzamanidis *et al.*, 2010). Specifically, overnight cultures of bacteria in biological triplicates were washed 3 times in 1x PBS. The OD_{600} of each was measured and the appropriate volume of the washed cultures was transferred to a glass tube to have a starting OD_{600} of 0.8 in 3 mL of 0.1 M KNO_3. Next, 1 mL of hexane was added in each tube and they were left still for 10 min at RT, followed by 2 min of constant vortexing. After the vortexing step, the tubes were incubated for 20 min at RT and then 100 μL from the level where the organic and the aqueous

phases were separated were transferred to a cuvette for an OD_{600} measurement. The percentage of hydrophobicity is calculated based on the decrease of the measured OD_{600} values using the $(1\text{-}A_1/A_0) \times 100$, where A_1 is the final OD_{600} value and A_0 is the starting OD_{600} value.

2.3.15.3 Microtiter Dish Biofilm Formation Assay

Assessment of the biofilm formation ability was performed for both bacteria monocultures and bacteria-yeast co-cultures as described by Merritt *et al.* (2005). In short, overnight cultures were diluted in either CDM46R- (bacteria monoculture) or CDM35R- (co-cultures) to an OD_{600} of 0.02 and 100 µL aliquots were transferred to 96-well polystyrene dishes. For each mono or co-culture, we used 4 replicates with the ratio between bacteria and yeast in the co-culture being 1:1. The dishes were incubated at RT for 48 hours, after which media and planktonic cells were removed and the wells were washed four times with ddH_2O. The adherent biofilm was stained with 200 µl of 0.01% [w/v] crystal violet for 15 min and subsequently the wells were washed four times with ddH_2O. Bound dye was solubilised with 200 µL of 80% ethanol-20% acetone, and the A_{570} was measured in a Tecan plate reader. Background absorbance for empty control wells was subtracted and the mean and standard deviation (SD) were calculated from the replicates.

2.3.16 Gas Chromatography - Mass Spectrometry (GC-MS) sample preparation, data acquisition and analysis

Analysis of extracellular and intracellular metabolites was performed on cultures that were sampled at mid-exponential phase. Briefly, 10 mL of culture were vacuum-filtered through nylon membrane filters (0.45 µm, Whatman™) and the filtered medium was immediately stored at -80 °C until further analysis. For sample preparation in 50 µL of filtered medium were added to 300 µL of cold (−20 °C) HPLC-grade methanol (Biosolve Chimie, France)/MilliQ water (1:1, [v/v]) were added and the samples were incubated at 72 °C for 15 min, while ribitol (Adonitol) (Alfa Aesar, UK) was added as an internal standard. The metabolites were centrifuged at 10,000 rpm and 0 °C for 10 min, and the supernatants were collected and dried with a vacuum concentrator.

The cell pellet on the membrane filters was washed in three rapid steps with 5 mL of PBS to ensure no contamination from extracellular metabolites. The polar metabolites were extracted by adding the washed cell-containing filter in 5 mL of cold (−20 °C) HPLC-grade methanol (Biosolve Chimie, France)/MilliQ water (1:1, [v/v]) and incubating for 1 h at −20 °C, followed by incubation at 72 °C for 15 min, while ribitol (adonitol, Alfa Aesar, UK) was

added as an internal standard. The mixture of metabolites and cell debris was centrifuged at 10,000 rpm and 0 °C for 10 min, and the supernatants were collected and dried with a vacuum concentrator.

Dried polar metabolites were derivatized with 40 µL of 20 mg/mL methoxyamine hydrochloride (Alfa Aesar, UK) solution in pyridine for 90 min at 37 ^{o}C, followed by reaction with 80 µL N-methyl-trimethylsilyl-trifluoroacetamide (MSTFA) (Alfa Aesar, UK) for 10 hours at RT. GC-MS analysis was performed using a Shimadzu TQ8040 GC-(triple quadrupole) MS system (Shimadzu Corp.) equipped with a 30m x 0.25 mm x 0.25 µm DB-50MS capillary column (Phenomenex, USA). 1 µl of sample was injected in split mode (split ratio 1:5) at 250 ^{o}C using helium as a carrier gas with a flow rate of 1 ml/min. GC oven temperature was held at 100 ^{o}C for 4 min followed by an increase to 320^{o}C with a rate of 10^{o}C/min, and a final constant temperature period at 320 ^{o}C for 11 min. The interface and the ion source were held at 280 ^{o}C and 230 ^{o}C, respectively. The detector was operated both in scanning mode recording in the range of 50-600 m/z, as well as in MRM mode for specified metabolites. The metabolite quantification was carried out by calculating the area under the curve (AUC) of the identified marker ions of each metabolite. Natural abundance isotopes were corrected using the Isotope Correction Toolbox (ICT)

2.3.17 Reconstruction of genome-scale metabolic models

Reconstruction of genome-scale metabolic models was performed with Carveme, based on Machado *et al.*, 2018 (Fast automated reconstruction of genome-scale metabolic models for microbial species and communities).

Contributions

- Dimitrios Konstantinidis (Structural and Computational Biology Unit, EMBL Heidelberg, Patil Group) identified the scientific question that lead to the invention of *Syn-A*, designed and conducted all the experiments, performed the data analysis, visualisation and interpretation, unless stated otherwise.
- Filipa Pereira (Structural and Computational Biology Unit, EMBL Heidelberg, Patil Group) provided feedback and discussion for data interpretation and optimization of the experimental design.
- Eleni Kafkia (Structural and Computational Biology Unit, EMBL Heidelberg, Patil Group) provided insights and protocols for GCMS sample preparation and data acquisition.
- Kristina Grkovska (Structural and Computational Biology Unit, EMBL Heidelberg, Patil Group) performed some transfers in the ALE experiments with *L. lactis* and *B. casei* and assisted with the cell surface trait experiments.
- Yongkyu Kim (Structural and Computational Biology Unit, EMBL Heidelberg, Patil Group) performed similarity analysis of the gene content of known vitamin producing strains and the Patil Group's collection of natural microbial isolates.
- Daniel Machado (Structural and Computational Biology Unit, EMBL Heidelberg, Patil Group) constructed the genome-scale metabolic models of the bacterial strains.
- Rose Gathungu (Metabolomics Core Facility, EMBL Heidelberg) performed the LCMS measurments for riboflavin and the nucleotides and did the initial data analysis.
- Rajna Hercog and Ferris Jung (Genomics Core Facility, EMBL Heidelberg) performed the whole genome sequencing.
- Diana Ordonez (Flow Cytometry Core Facility, EMBL Heidelberg) helped with setting up the protocol for the FACs analysis.
- Mandy Rettel and Frank Stein (Proteomics Core Facility, EMBL Heidelberg) performed the proteomics sample preparation and initial data analysis.
- Kiran Raosaheb Patil (Structural and Computational Biology Unit, EMBL Heidelberg, Patil Group) supervised the project, designed experiments and helped with data interpretation.

Chapter 3: Model guided evolution of costly metabolic phenotypes

3.1 Introduction

Industrial microbial fermentations largely rely on the selection of appropriate starter cultures, in order to retain the fermentation efficiency and the same quality of the end-product constant, without batch to batch variations. The regulations concerning use of genetically modified organisms (GMOs) has led producers to select high-performing strains from banks of natural or artificial variability (Steensels *et al.*, 2014). It is possible to obtain artificial variability by using random mutagenesis techniques or mass-mating, however it is difficult to accurately predict the results of these methods trying to acquire specific phenotypic characteristics.

Adaptive laboratory evolution (ALE) is one of the popular alternatives and is often used with the purpose to enrich the number of beneficial mutations and subsequently improve the desired phenotypic characteristics of an organism. Nevertheless, cell fitness is rarely aligned with the desired characteristic and as a result enrichment of mutations through ALE will not improve the application performance. Towards this direction in this thesis we describe how our novel algorithm which uses genome-scale metabolic model simulations can lead to the design of environmental conditions that will create selection pressure for the production of a target compound. Then ALE can be performed in the model designed environmental conditions to straightforwardly select for improved production of the target compound. In this first chapter, we introduce the necessary information that led to the creation of the novel algorithm *EvolveX* and the basic assumptions that are needed for the use of the algorithm. Finally, we describe briefly the target compounds and their importance, aiming to justify why we chose their production as proof of concept for the applicability of *EvolveX*.

3.1.1 Adaptability and exaptations

In the field of evolutionary biology, the term adaptation is used to describe the phenotypic characteristics of one organism that were selected for a specific role (Brosius, 2019). These features are the ones most commonly used by scientists, to explain how natural selection acts upon organisms. Another important concept in evolution is the term exaptation, also known as preadaptation, which is used to describe neutrally evolving features that don't seem to offer an immediate advantage to the organism, or serve a specific function. However, traits that evolve as exaptations may at some point of evolution offer an advantage that is unrelated to their origin (Barve and Wagner, 2013). Numerous prominent examples of exaptations have been described in the animal kingdom, as early as Darwin and his contemporaries. For example, Wallace (Wallace, 1889) described the different uses that tails may exhibit, such as enhancing balance, being used for climbing, or even as a temperature

regulator. Similarly, the front limbs adapted for flight in bats and the wings of penguins are used as flippers (Brosius, 2019).

Apart from the above-mentioned examples of exaptation that involve the novel function of a previously developed characteristic, researchers expanded the concept of exaptation to the molecular level. The most commonly described example is the duplication of a functional gene, where one copy retains the original function, while the second one acquires a novel or similar action as the original (Brosius, 2019). The numerous examples of exaptation that have been described over time were proof of the importance of this kind of adaptation, however, scientists needed a good system to consistently study exaptations both at the genotypic and phenotypic level and metabolism proved to be such a system. In general, the metabolism of the cell consists of a large number of chemical reactions which occur through the action of various enzymes aiming to synthesise all of the necessary compounds for survival and increase of biomass (Barve and Wagner, 2013). Metabolic reactions are dependent on molecules that are available in the environment and since environmental conditions often fluctuate, cells need to have an easily adaptable metabolism, to survive.

Microorganisms are a good example to showcase the adaptability of metabolism, considering the fact that they are affected greatly by environmental conditions and still they manage to thrive in all possible niches. Part of their success to inhabit dynamic environments is owed to their variable metabolic networks which depend on complex gene regulation that determines which reactions will be active or not (Sambamoorthy *et al.*, 2019). This observed adaptability is partially acquired through exaptations and researches are taking advantage of it for various applications. The best characterised microbial exaptation is the ability to proliferate when different carbon sources are available. The addition of a single extra reaction to a network may lead to the emergence of a novel complex phenotype, that can eventually prove useful. For example, *Pseudomonas putida* that was selectively evolved to consume xylose expressed the ability to utilise arabinose as well, even though the parental strain could not use this carbon source (Barve and Wagner, 2013). Therefore, it is clear that an initial period of pre-adaptation can be used as a useful tool, especially for cases such as bioremediation of complex compounds. In these cases, researchers take advantage of laboratory evolution, aiming to adapt a microbial species or a microbial community to specific dangerous chemical pollutants that need to be removed from the environment. The use of long or short term laboratory evolution allows both the study of the adaptation mechanism and the improvement of bioremediation rate (Poursat *et al.*, 2019).

3.1.2 The challenge of predicting the outcome of natural selection

The majority of evolutionary studies involve the simultaneous use of many identical initial strains or populations which undergo the evolutionary procedure in parallel. Even though these populations evolve independently, they converge towards the same phenotype and similar genetic mutations and this phenomenon is characterised as parallelism. Parallelism is observed in every kind of experimental evolution approach and for all the various species that are used in these experiments, thus it is safe to assume that it is driven by natural selection. Nevertheless, the most important outcome of parallelism is the rise of the hypothesis that the evolution process could be predicted (McDonanld, 2019). Evolutionary biology is traditionally considered a more descriptive than a predictive field, as natural selection can often act in an unpredictable manner, which cannot be anticipated only by studying past evolutionary events, due to the effect of random processes like genetic drift (Nosil *et al.*, 2018). Moreover, accurate predictions are tough to make even in ALE experiments where usually a single unicellular organism is grown under strictly monitored environmental conditions. The reason for this is the observation that these small organisms exhibit the ability to diverge during the evolution process and create smaller subpopulations specialised for the use of resources that are available in smaller niches. These niches are usually not created by the experimental design, but commonly by the release of metabolic by-products or the rate at which initial nutrients are being consumed. These subtle environmental changes are adequate to initiate a different evolutionary pressure and lead to the emergence of subpopulations that evolve different characteristics than expected. Moreover, it is possible that these subpopulations start to compete, a phenomenon known as clonal interference, which may lead to one non-predicted population outcompeting all the others (McDonanld, 2019). It is clear that the predictability of evolution and the ability to successfully direct it towards a desired outcome remains an unachieved aim in research, which could potentially be a step closer to accomplishment by focusing on the creation of more specific environmental conditions. Such conditions need to take into consideration the metabolic capabilities and the genetic background of an organism, apart from solely direct pressures, such as temperature and nutrient resources.

3.1.3 Genome scale metabolic models

The advances in systems biology and high-throughput experimental methods have led to a deeper understanding of the biochemical and genetic background that shape cellular functions. Through mathematical functions, it is possible to represent this information as a network to build models that allow the computation of the genotype-phenotype relationship

(Lewis *et al.*, 2012) and to computationally analyse the metabolism, gene regulation and other cellular functions (O'Brien *et al.*, 2015). The models that were developed to shed light in the complexity of the genotype-phenotype relationship are the genome scale metabolic models (GEMs) which aim to reconstruct the complete metabolism of one organism with the use of all the known metabolic reactions and metabolic genes described for a specific organism. The first GEM was constructed for *Haemophilus influenza*, shortly after its genome was sequenced (O'Brien *et al.*, 2015), while the first GEM for an eukaryotic organism was constructed in 2003 for *S. cerevisiae* and included three cellular compartments, the cytosol, the mitochondria and the extracellular space (Österlund *et al.*, 2012). The predictive power of GEMs has improved over the past years, allowing scientists to evaluate how gene knockouts or changes in nutrition potentially affect microbial cells and thus are widely applied for the rational design of industrially used strains, discovery on novel drugs against pathogens or to answer fundamental questions about species interactions in microbial communities (Machado *et al.*, 2018).

The most popular approach for constructing GEMs is the Constraints-Based Reconstruction and Analysis (COBRA, Heirendt *et al*. 2019) approach which accounts for the constraints on the phenotype of an organism that are imposed by the physiochemical laws and the genome of the organism. The use of constraints is necessary, since metabolism is a complex network of biochemical reactions which are regulated by the availability of the required substrates, the activity of the enzymes that catalyse them and the laws of thermodynamics. Therefore, it is understood that a realistic model needs to account for the existence of substrates in the environment of the cell, the enzymatic capabilities of this cell as well as the stoichiometry of these reactions (Lewis *et al.*, 2012). These constraints create the dimensional space of possible solutions where each solution is a potential phenotype and thus application of the constraints aims to limit the number of solutions and keep the computational complexity at a reasonable level. A commonly used tool of COBRA is Flux Balance Analysis (FBA) which is used for the analysis of the metabolites' flow through the metabolic network. The most important assumptions of FBA are that growth follows a steady state and that all mass that enters the system must also leave (mass balance analysis). The use of FBA aims to identify the best solution from all the solutions that form the 'solution space' according to the initially described parameters. It is a simple tool that can be used broadly, since its use requires information only for the stoichiometry of metabolic reactions and the mass balance in hypothetical steady state conditions. However, the use of FBA does not allow the prediction of metabolite concentrations and does not take into consideration the

regulatory aspects of metabolism, such as enzyme activation or gene expression (O'Brien *et al.*, 2015). This type of modelling became one of the most important tools in biotechnology leading to the *in silico* design of microbial strains that perform specific bioconversions for the production of fuels, chemicals or materials from renewable resources. The importance of GEMs is reflected on the fact that numerous patent applications that were filled for industrially important microorganisms, mention the use of those models (Lopes and Rocha, 2017).

The predictive power of models has been proven to be a useful tool for evolutionary biologists. For example, evolutionary optimality models have been used to predict potential changes that occur during adaptation and estimate changes in virulence or enzymatic activity (Harcombe *et al.*, 2013). Another field of evolutionary biology that can benefit from GEMs is the elucidation of evolutionary relationships between organisms. Traditionally the untangling of such relationships is based on the comparison of conserved sequences of genes and similarities in protein function, but with the use of GEMs, scientists can compare whole metabolic pathways and get a better insight into the process of evolution. Furthermore, the function and emergence of paralog genes and enzymatic redundancy can be studied either within one single species or among diverse organisms (Österlund *et al.*, 2012). As it was mentioned in a previous chapter, the result of ALE experiments is closely dependent on the choice of the selection pressure and the establishment of the proper environmental conditions can often prove complex and problematic. Recently, a model-based algorithm that could be used to optimise ALE experiments was developed, aiming to offer standardisation in the design and execution of the evolutionary procedure (LaCroix *et al.*, 2017). However, it remains challenging to perform ALE experiments that aim to select for traits that are not associated with cell fitness. In the next chapter I will describe the design of a novel algorithm, termed *EvolveX* which can be used to tackle this problem.

3.1.4 *EvolveX*: A novel way to evolve non-fitness associated traits

The novel algorithm *EvolveX* was developed in the Patil group, with the aim of identifying the evolution conditions that will lead to the enhanced production of costly metabolites. These metabolites are not directly connected to cell fitness under the normal environmental conditions that the organisms encounter. *EvolveX* is used to computationally design a chemical environment which in combination with genome-scale metabolic models develops a selection pressure on the targeted metabolic fluxes, meaning that a dependency is established between cellular fitness and the production of the target metabolites. The designed environment which is defined as the evolution niche is used to perform an ALE

experiment where the straightforward selection based on growth leads to the accumulation of mutations and adaptations which will allow them to keep the selected phenotype when the evolved cells return to their naturally occurring environment, also termed as the target niche (Figure 3.1) (Jouhten*, Konstantinidis*, *et al.*, manuscript in preparation).

In its basis, *EvolveX* is aiming to predict the translation of the interplay between genotype and phenotype, using genome-scale metabolic models. These models were extended for identifying the underpinnings of a complex metabolic trait, by pinpointing the metabolic fluxes that give rise to the desired complex metabolic trait. A minimum set of such fluxes that need to be up- or down-regulated are defined in *EvolveX* as a mixed-integer linear programming (MILP) problem which can be used to predict mutant metabolic phenotypes (Shlomi *et al.* 2005). This calculation can be used to predict whether a flux needs to be increased or decreased, so that the concentration of the target compound increases. During all the calculations, the minimum and maximum fluxes of the wild-type phenotype are taken into consideration, thus creating a group of possible solutions.

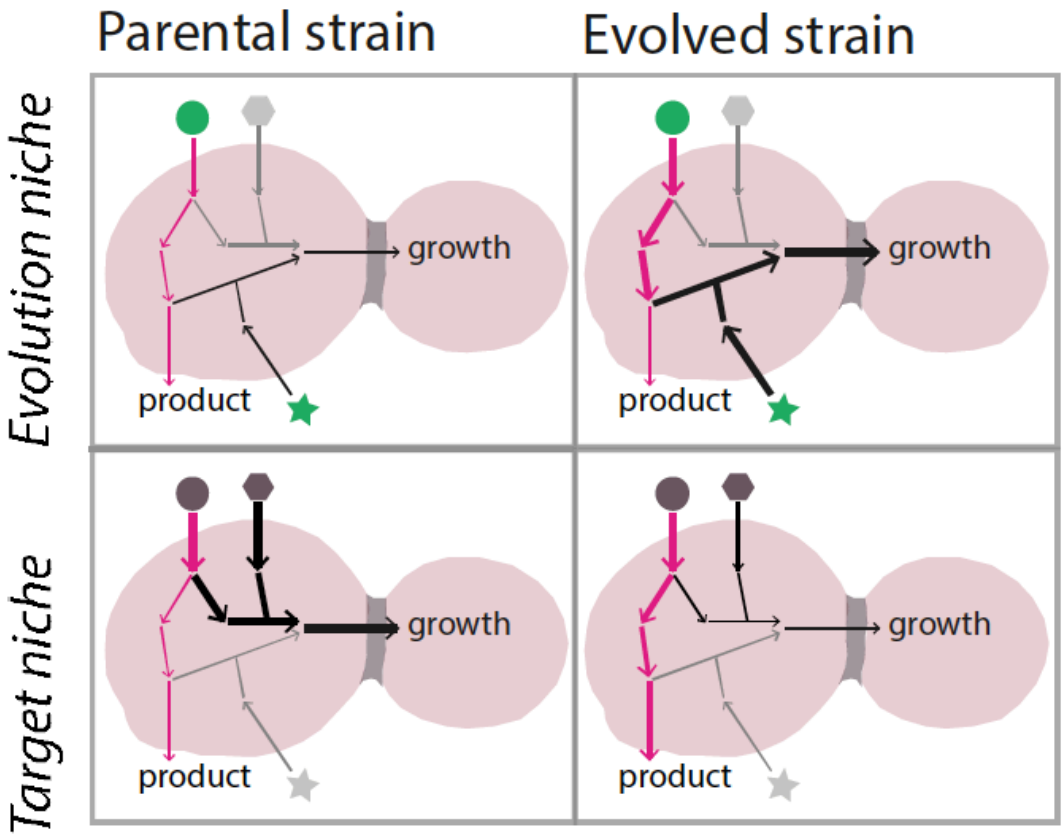

Figure 3.1. Methodology of applying *EvolveX*. Illustration of the parental and an evolved strain in the Evolution niche designed with *EvolveX* and in the Target niche, used for a biotechnological application. Red arrows represent up-regulation of target metabolic fluxes, grey arrows absent fluxes and thick arrows represent increase of metabolic fluxes.

3.1.5 Creation of Evolution niches with defined chemical composition

The trait definition as a set of fluxes requiring up- or down-regulation allows to identify environmental transformations, the *Evolution niches*, where regulation of the requirements become directly connected to fitness. The environmental conditions of the *Evolution niche* include nutrients (e.g. carbon and nitrogen sources), inhibitors (e.g. substrate analogues) and regulatory triggers (e.g. 2-deoxyglucose) all of which are available and can be taken up by the cells. It is important to initially determine the optimal yield of biomass when the nutrients in the *Evolution niche* are being used under the effects of available inhibitors and metabolic effectors. The strength of the selection pressure under the conditions of the *Evolution niche* can also be predicted by determining the minimum possible fitness coupled with the up-regulated targets and similarly the potentially maximum for down-regulated targets. The total selection pressure exposed on the flux re-regulation targets by the particular *Evolution niche* was defined as the sum of the worst-case flux couplings to growth. Subsequently after evaluating the strength of the selection pressure on the selected targets, the minimum number of re-regulated fluxes that the strongest selection pressure acts on, in the particular *Evolution niche* was estimated. The final score which was used to estimate whether an Evolution niche is appropriate for the adaptive evolution of a desired complex metabolic trait, was based on the strength of the selection pressure on the flux re-regulation targets, the coverage of the flux re-regulation targets by selection pressure stronger than in common laboratory conditions, and the number of chemical components in the *Evolution niche*. The number of chemical components in the *Evolution niche* was considered in the score matrix as the uncertainty of combining effects to the selected components that was expected to increase as their number in the *Evolution niche* increases.

3.1.6 Wine: a complex product formed synergistically by biotic and abiotic factors

Wine is an alcoholic beverage that has been closely associated with the development of human civilisation, since its first steps and the beginning of agriculture. The first signs of wine making have been identified in a pottery jar, dating back to the Neolithic period and evidence of fermentation for the production of alcoholic drinks expand from China to Egypt and Europe, even to the New World (Marsit and Dequin, 2015). In modern day society, wine still remains one of the most popular drinks which influences numerous social and cultural aspects of human life. One of the reasons why wine is so popular is the diversity in flavour and aroma, which are termed as *terroir* and are shaped by diverse factors such as climate, soil and the variety of the grapes (Belda *et al.*, 2017). Another essential aspect in winemaking is the microorganisms that are used during the fermentation process, with the most important

one being *S. cerevisiae* which can grow under high concentrations of sugars and ethanol. Traditionally the oenological characteristics of wine are improved through the selection of different yeast strains. These strains are most commonly wild strains which are transferred to the lab and their phenotypic characteristics are evaluated, but this procedure is time consuming and expensive (Garcia *et al.*, 2012).

Apart from the role of promoting the fermentation process, yeasts are also actively shaping the flavour and aroma characteristics either by the production of ethanol which acts as a solvent and extracts flavour components from the grapes or by producing various secondary metabolites with specific aroma properties. Mainly higher alcohols and esters that are formed during the alcoholic fermentation influence the sensory properties of the resulting wine (Gonzalez and Morales, 2017). These organic compounds are known as fusel alcohols, deriving from the German word *Fusel* (bad liquor), as high amounts of these alcohols are obtained during the distillation of spirits and are associated with off-flavours in the final product. However, fusel alcohols are desired in small quantities and it is known that they originate from the metabolism of amino acids through a group of reactions known as the Ehrlich pathway (Hazelwood *et al.*, 2008). Amino acids are freely available in the wine must and are taken up slowly during the fermentation process. The amino acids that can be assimilated to the Ehrlich pathway are the following: valine, leucine, isoleucine, methionine (Figure 3.2A) and phenylalanine (Figure 3.2B). They are initially transaminated and subsequently decarboxylated and into fusel aldehydes, which finally are reduced into alcohols (Belda *et al.*, 2017). Three of the most important compounds originate from the metabolism of the branched-chain amino acids (leucine, valine, isoleucine) and they are the isoamyl alcohol (3-methylbutanol), the isobutanol (2-methylpropanol) and the active amyl alchol (2-methylbutanol). Other important aromatic compounds are produced when phenylalanine enters the Ehrlich pathway leading to the production of phenylethanol and phenylacetate (Hazelwood *et al.*, 2008, Belda *et al.*, 2017).

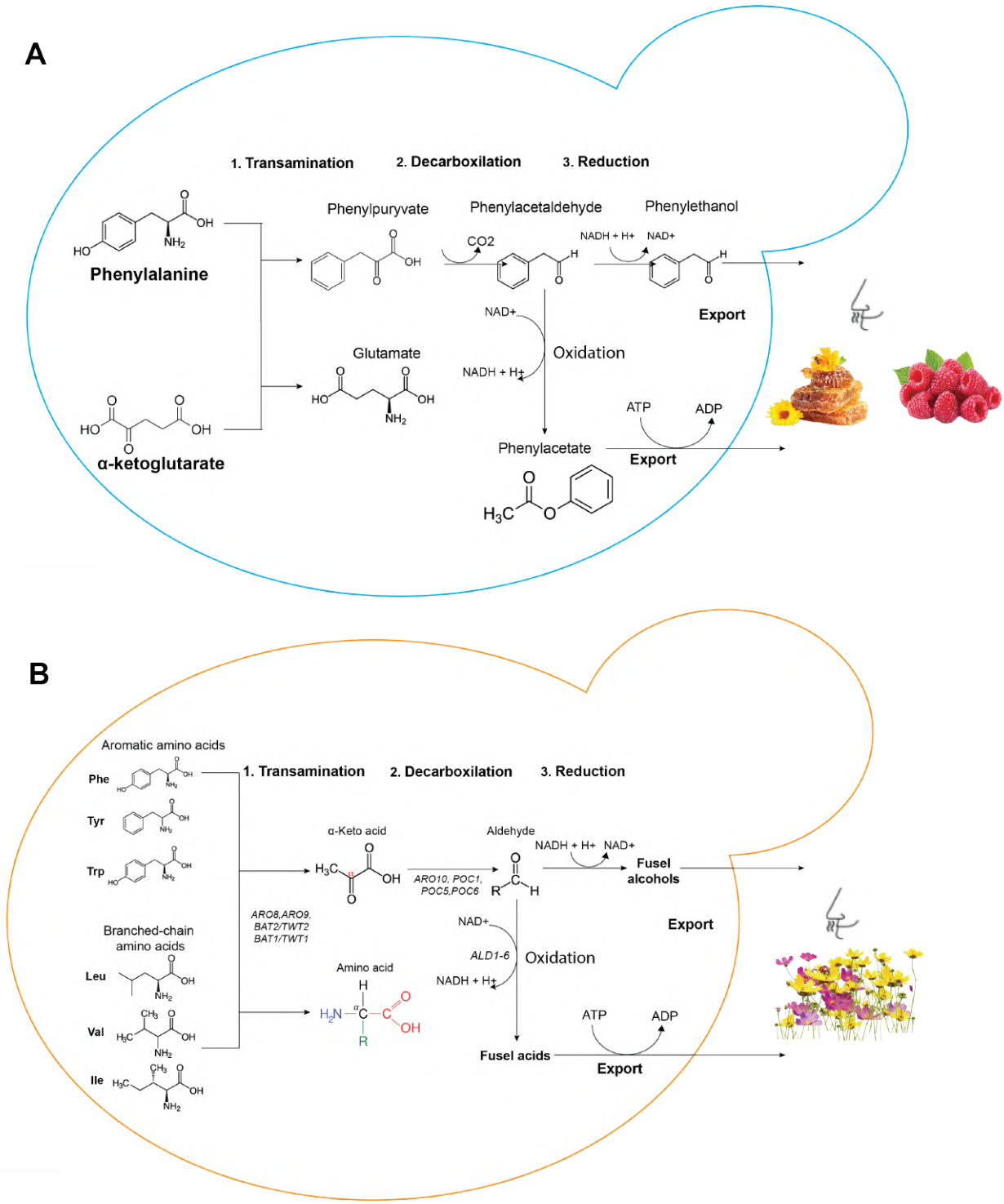

Figure 3.2. The Ehrlich pathway. (A) Catabolism of phenylalanine via this pathway leads to the formation of phenylethanol and phnylacetate, which can be exported from the cells, contributing to honey or raspberry like aroma and flavour. (B) Aromatic amino acids and branched-chain amino acids are catabolized through the Ehrlich pathway leading to the formation of fusel acids and alcohols as the final product. Secretion of these compounds is associated with a floral aroma in wine (Adapted from Belda *et al.*, 2017).

3.2 Results and discussion

Researchers aiming to improve specific phenotypic traits of microorganisms, but due to various reasons (Burgess *et al.*, 2006) they are frequently unable to use methods of genetic engineering. In cases where the desired traits have complex genetic background use of ALE offers an interesting solution. However, up to now, the use of ALE is limited solely to obtain traits that increase cellular fitness. In this chapter we show that specific, not necessarily fitness aligned metabolic traits can be specifically evolved by computationally designing the chemical environment that will be used for ALE. We used genome-scale metabolic models to develop a dependency between this designed environment and specific metabolic fluxes that lead to the production of a target compound. As a proof of concept, we demonstrate the validity of this approach by evolving new metabolic traits in *Saccharomyces cerevisiae* for food applications where genetically modified organisms (GMO) are undesired.

3.2.1 Creation of evolution niches to test as proof of concept

The effectiveness of *EvolveX* under real conditions needed to be assessed and for this reason we set out to create the proper experimental conditions that would put it to the test. As a case study, we decided to evolve strains of wine yeast strains for the increased production of aroma compounds which are associated with a pleasant bouquet in wine. A large variety of such compounds are produced during the fermentation process, so we decided to focus on increasing the production of two separate groups through separate evolution experiments. The first set of target compounds consisted of phenylethanol and the subsequently derived acetate ester, namely phenylethyacetate which gives to wine a rose and honey scent as well as a raspberry-like flavour. The second set of compounds included two higher alcohols (2-methyl-1-butanol and 3-methyl-1-butanol) and their acetate esters (2-methylbutylacetate and 3-methylbutylacetate). The synthesis of these compounds derives from branched-chain amino acids and has been reported to contribute to the floral aroma of wine. Before the evolution experiment could start, the appropriate evolution niches had to be computed by *EvolveX*, by calculating the minimum number of reactions required for the regulation of the fluxes of the wild-type strain that would allow the optimal synthesis of the target compounds. Secondly, these reactions were used to form a scoring matrix of carbon and nitrogen sources and the highest scoring combinations were used to synthesise the evolution niche for each of the two cases, with each niche being named by the carbon source that it contained. For the increased production of phenylethyl acetate the evolution niche contained glycerol as the sole carbon source, and as nitrogen sources L-threonine and L-phenylalanine, the direct precursor of the target compounds (Evolution niche G). None of the

direct precursors of branched-chain amino acid-derivatives higher alcohols were included in the second evolution niche, which contained ethanol as carbon source and L-arginine and glycine as nitrogen sources (Evolution niche E).

3.2.2 Adaptive evolution for the increased generation of aroma compounds

After the computation of the evolution niches was completed we moved on to select the *S. cerevisiae* wine strains that would undergo the evolution process. Initially, twelve different industrial strains were tested for their ability to grow in both niche G and niche E. Most of the strains were able to grow efficiently in *Evolution nich*e E after 4 days of culture. On the other hand, only four strains exhibited the ability to grow in *Evolution niche* G even though they were cultured for a period of seven days. One of the strains, namely QA23 was able to grow in both evolution niches and thus we decided to perform both evolution experiments with this strain. Strain QA23 is a popular strain amongst wine makers, as it is known for its low nutrient and oxygen requirements (Borneman *et al.*, 2011) and can be commercially used for the production of a variety of wine styles, such as Chardonnay, Sauvignon Blanc and Gewürztraminer. Three individual colonies of our parental strain were inoculated in both of the evolution niches and each of the six cultures underwent a serial transfer experiment. These transfers where performed when growth was equal or higher than an OD_{600} value of 3 and the ALE procedure lasted for a total of 2 months. During this period, each culture evolved into a separate population, but the total number of accumulated generations in each evolution niche was very similar for all the triplicates. In more detail, the populations evolved in niche E were transferred 17 times with 107 generations passing on average, while the G evolution niche populations were transferred 19 times and an average of 100 generations passed. Initial fitness in both evolution niches was low, but after the 5 first initial transfers, the fitness increased enough to reduce the transfer volume by half. Further adaptation led to shorter time periods between transfers. In more detail, initially cultures in evolution niche E were transferred every 96 h, with the transfer time being reduced to 72 h after 10 transfers and even further to 48 h after 14 transfers. Similarly, the transfer time in evolution niche G was reduced from 120 h to 96 after 8 transfers, and further to 72 h after 11 transfers. As a control experiment the selected strain was transferred for the same time period in glucose minimal medium and the number of transfers was almost double (30 transfers). This higher number of transfers in the glucose medium shows that growing to the conditions of the evolution niches is challenging for the yeast cells and requires a lot of

metabolic adaptations. We also did not observe any increase or decrease in the optical density of the glucose cultures, since our parental strain could originally grow efficiently under these conditions and therefore no direct environmental pressure was acting on the cells. All throughout the duration of the experiments, glycerol stocks of intermediate populations were prepared for future reference, as well as for the final populations. The next step after the completion of the serial transfer experiment was to evaluate whether the improvement in growth that we detected correlated with improved phenotypic characteristics, as we had initially hypothesised.

3.2.3 Phenotypic characterization of the evolved isolates

After the evolution had finished we wanted to investigate the differences that had occurred in evolved strains' phenotype. According to the predictions of *EvolveX*, production of the target compounds is linked to fitness in the evolution niche and therefore we chose to further characterise the clones that exhibited the best growth under the ALE conditions (Figure 3.3). To isolate single strains, we started by platting the final populations on wine must mimicking (WMM) medium agar plates, a harsh environment due to its low pH value and the high sugar content (Marsit and Dequin, 2015). This medium was the target niche of our experiment and hence clones that lost the ability to grow under such stresses were of no interest to us. All the populations were able to grow under these conditions and the single strains that were isolated were inoculated once more in WMM to confirm that they do not exhibit any decrease in fitness.

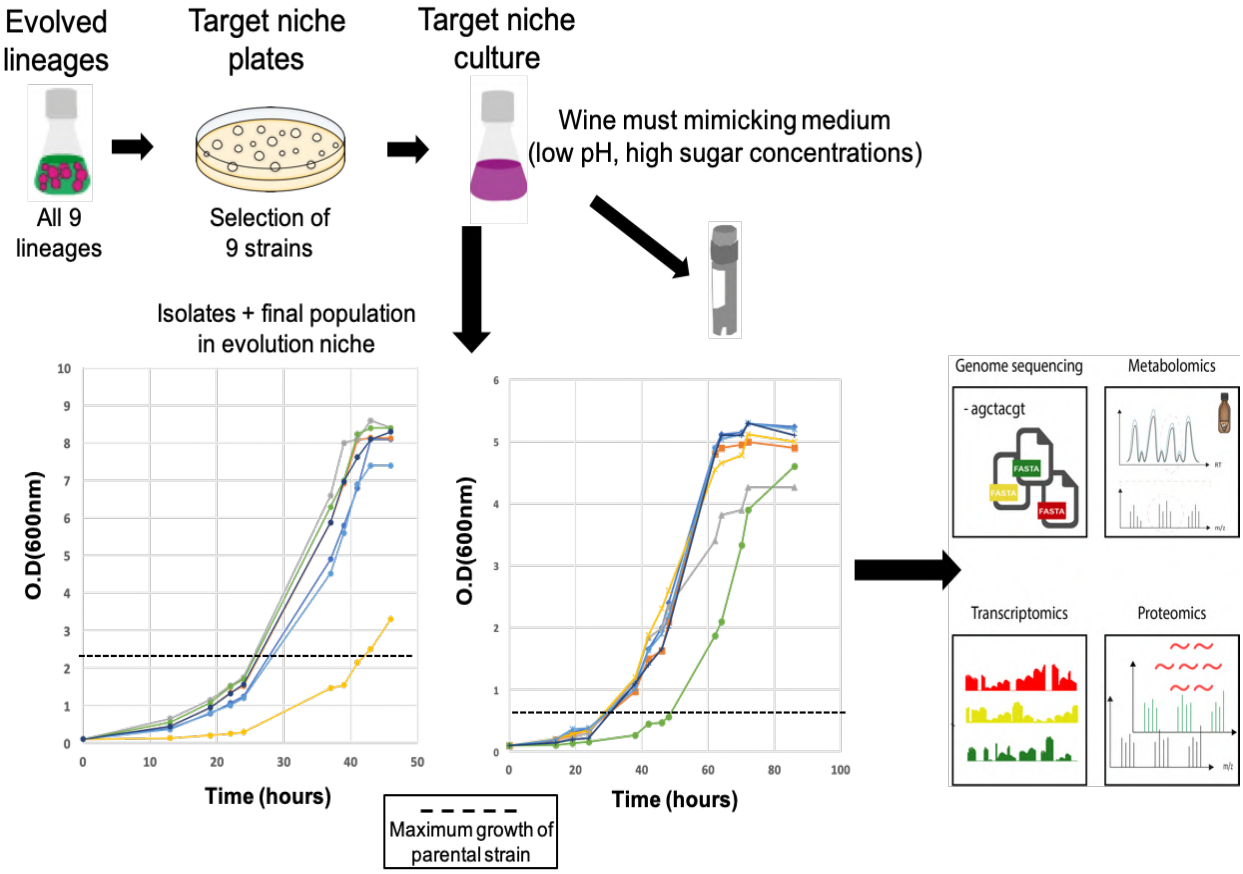

Figure 3.3. Experimental design for the increased production of aroma compounds. In the Evolution niches designed with *EvolveX*. After completion of the serial transfer experiment the final population were transferred to the target niche (WMM) and single strains were isolated. These isolates were tested for their ability to grow without defects in the target niche and subsequently were stored as glycerol stocks and transferred back to the evolution niche. Their fitness in the evolution niche was tested and the strains that exhibited the best growth kinetics were chosen for further characterisation.

From each evolved population nine strains were isolated and brought back to their particular evolution niche for estimation of their growth kinetics. Introduction to the wine must mimicking medium before the growth kinetics in the evolution niches was expected to separate strains that exhibited adaptation phenomena from real mutations. Each set of isolated strains was compared with the corresponding final population (Figures 3.4 and 3.5) and one isolated strain out of each evolution lineage performing comparably well to the corresponding final population was selected for phenotyping.

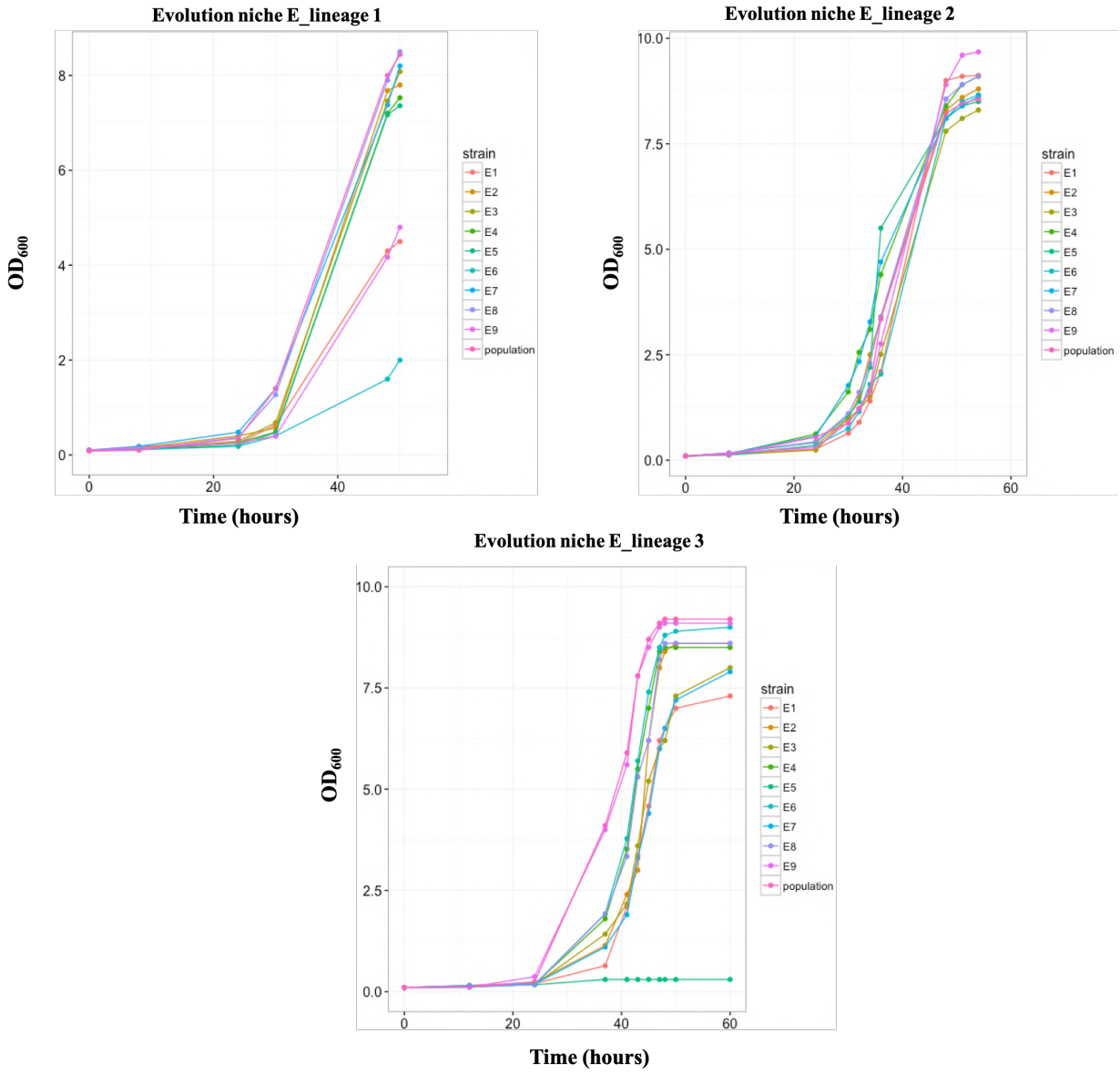

Figure 3.4. Growth kinetics of isolated evolved strains in the evolution niche E. From each lineage nine evolved strains were isolated and their growth was compared to the growth of the whole final population. Each panel, represents one of the three final populations of the ALE in evolution niche E.

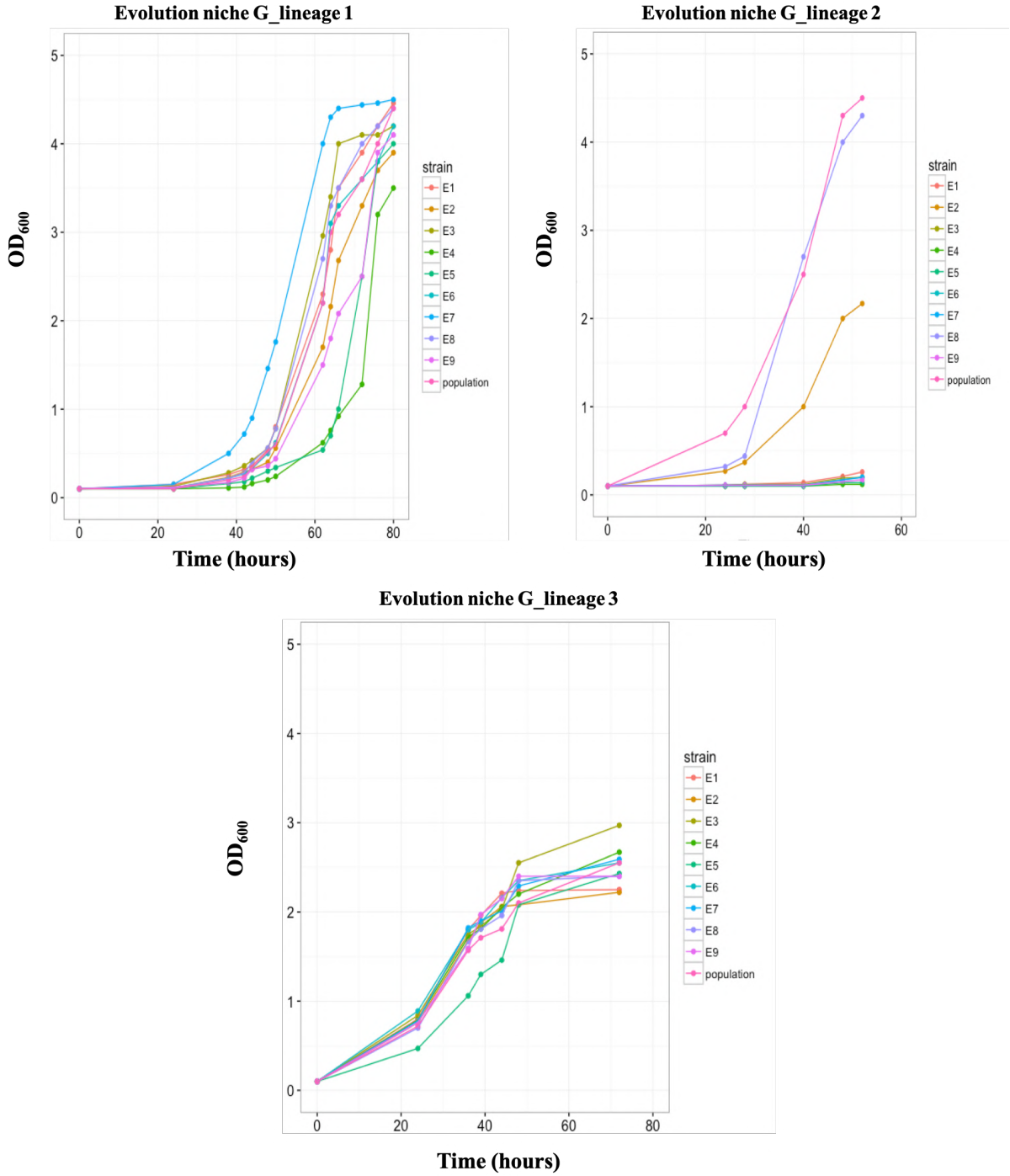

Figure 3.5. Growth kinetics of isolated evolved strains in the evolution niche G. From each lineage nine evolved strains were isolated and their growth was compared to the growth of the whole final population. Each panel, represents one of the three final populations of the ALE in evolution niche G.

In collaboration with Dr. Ramon Gonzalez (Instituto de Ciencias de la Vid y del Vino, Logrono, Spain) we investigated the ability of the strains to produce aroma compounds during the fermentation of synthetic wine must (SWM). The selected strains were able to produce higher amounts of the target compounds compared to the parental strain (Figure 3.6), while retaining the same fermentation performance (Figure 3.7). In more detail, the strain that was isolated from the third lineage of ALE in niche G (G3), produced statistically

significant higher amounts of phenylethyl acetate, with an increase of a log_2fold change equal to 0.75 compared to the parental strain (Figure 3.6A). The two strains from niche G, produced either smaller or slightly higher amount of aroma compounds originating from branched-chain amino acids, in comparison to the parental strain. From the two strains that were evolved in niche E, the best performing one was the strain isolated from the second lineage (Figure 3.6B). Interestingly this strain produced slightly less amount of phenylethylacetate compared to the parental strain. However, neither of the strains exhibited a significantly improved phenotype for the target compound 2-methyl-1-butanol.

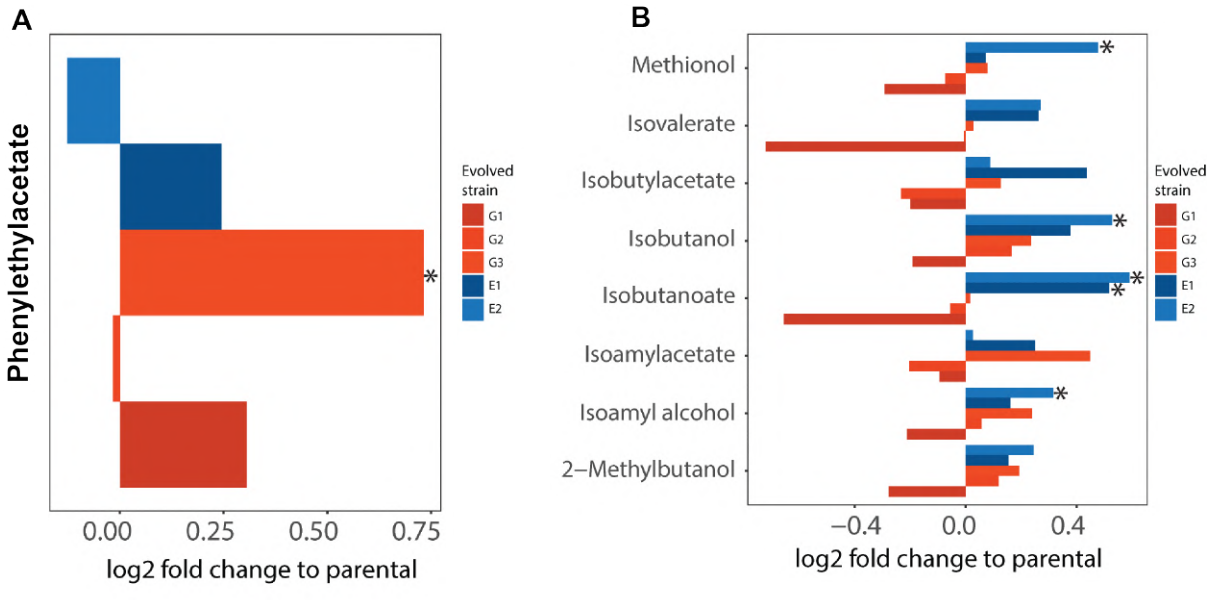

Figure 3.6. Production of volatiles by isolated evolved strains. (A) Production of phenylethyl acetate by strains evolved in both evolution niches. (B) Production levels of branched amino acid derived aroma compounds by strains evolved in both evolution niches. In both cases, strains evolved for a specific trait exhibit a stronger phenotype. All productions levels are calculated as log_2fold changes and compared to the parental with Dunnett test ($p < 0.05$).

Apart from the production of aroma compounds, also the general fermentation capabilities of all the strains were evaluated. Especially, the two strains that exhibited the best aroma compounds producing phenotype, namely E2 and G3, had the same yield as the parental for fermentation related metabolites, such as ethanol, glycerol and acetate (Figure 3.7). The levels of residual sugars were a bit higher, but still comparable with the detected levels of the parental strain. These data suggest that even after the evolution process the evolved strains retain the same fermentation phenotype as the parental strain and thus are potentially ready to be used in large-scale industrial settings.

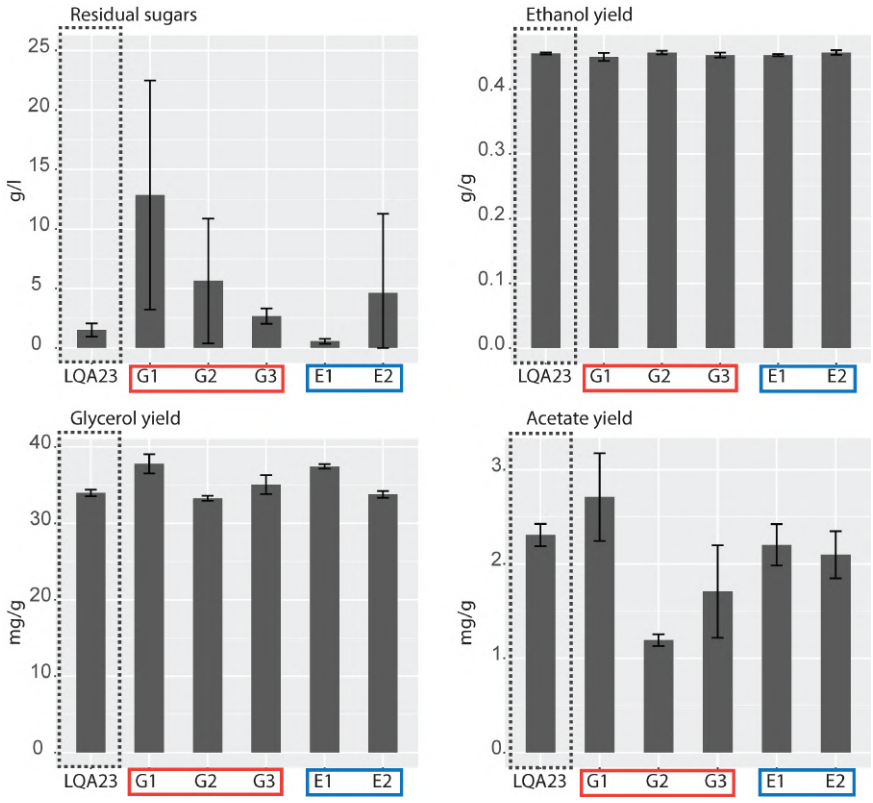

Figure 3.7. Evaluation of the fermentation phenotype. The fermentation profiles of the evolved strains remain comparable to their parental strain (LQA23) and hence they retain the wine fermentation phenotype. Strains evolved in the evolution niche G are labelled as G1, G2 and G3, while strains evolved in evolution niche E as E1 and E2.

Even though the results from these small-scale fermentations were encouraging, our target compounds are highly volatile and are being produced in relatively small amounts by the cells. Moreover, wine fermentation is highly dynamic and can potentially be effected by the substrate, meaning the natural variability between different batches of wine must. It has been observed that certain phenotypic characteristics are absent during the fermentation of SWM. These facts in combination with the fermentation process being performed semi-aerobically led to a reproducibility issue when the same strains were used to repeat the fermentations. For this reason, we decided to further continue the evolution process, aiming to acquire a strain expressing the desired phenotype more strongly.

3.2.4 Continuation of the ALE process enhances the observed phenotype

The strains that were isolated from evolution niche E and exhibited an improved phenotype during fermentation were subjected to a second round of ALE for a total duration of 1 month and 15 more transfers which equalled to an average of 97 generations. In the case of evolution niche G, we decided to continue the ALE process starting with 2 of the final populations, instead of single strains. The duration of the second round of ALE was also 1

month and the number of generations on average equalled to 65 in a total of 13 transfers. Similarly to the previously described methodology the 4 new final populations were transferred to solid WMM and 6 clones were isolated from each one of them and their fitness was evaluated in the appropriate evolution niche. The best performing strains were once again characterised for their ability to produce the target compounds (Figure 3.8). To avoid the reproducibility problems that were previously observed, this time the fermentations were performed under a variety of conditions using always the same natural wine must (NWM), a fact that limited the number of strains that could be used. It is worth mentioning that when the fermentations were performed in SWM the amount of phenylethyl acetate that was detected was practically the same for both the parental and the evolved strain, while fermentation of NWM led to a 1.5-fold increased production by the evolved strain, so we continued our analysis solely with the anaerobic NWM fermentations. Under these specific conditions the strain isolated from the evolution niche E produced the same amount of phenylethyl acetate as the parental and 1.2-fold increased amount of the target aroma compounds 2-methyl-1-butanol and 3-methyl-1-butanol, which are being quantified together with the metabolomics method that we used (Figure 3.8A). These data suggest that the evolutionary procedure in the *EvolveX* designed environments is directed for specific compounds, even though, we observed increased production of compounds by both evolved strains compared to their parental. Such convergence of the evolved phenotypes was expected, since the selected evolution niches were overlapping in respect to the up-regulation of the fluxes that lead to the production of phenylethanol. The large pool of aroma related compounds that was detected during the fermentations allowed for a better visualization of the relationship between the evolved strains and their parental through a principal component analysis (Figure 3.8B). Indeed, according to PC1 and PC2 both evolved strains have different directions compared to the parental strain, while PC2 can be used to separate the evolved strains according to their evolution niche. All the above, suggest that ALE in each designed niche leads to a distinct “aroma phenotype” during the fermentation process.

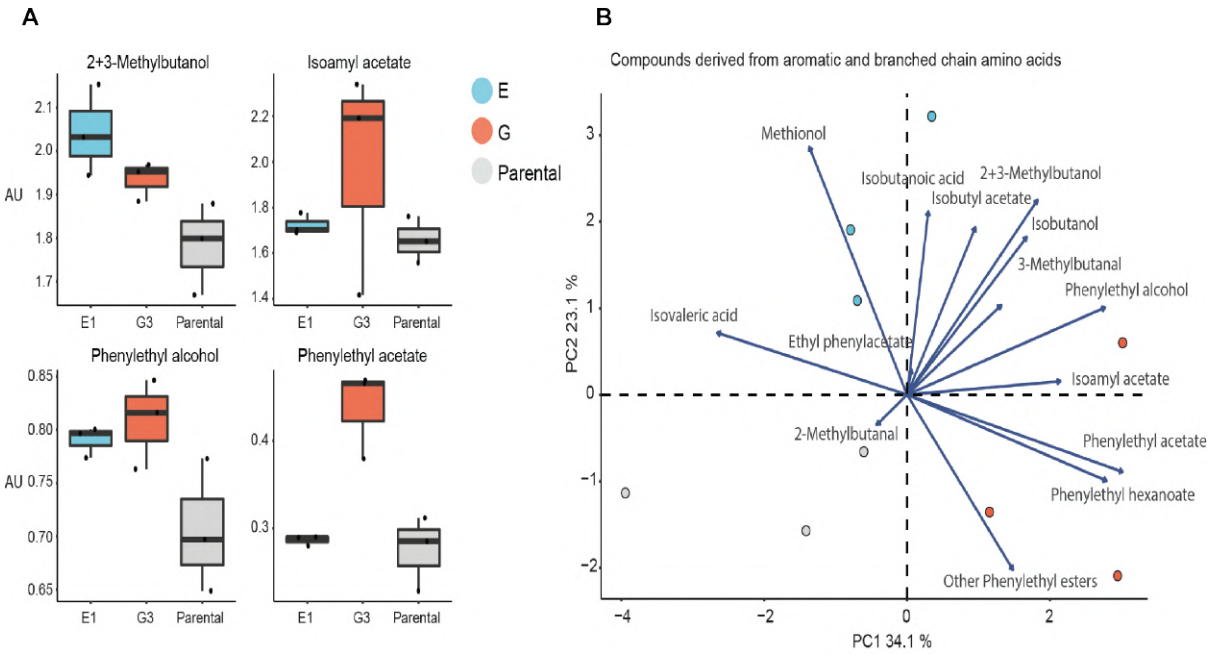

Figure 3.8. Aroma compound analysis of natural wine must fermentations of parental and evolved strains revealed targeted improvements. (A) Target compound abundances obtained in natural wine must fermentations of the evolved strains (E2: light blue; G3: orange) relative to the parental strain (grey). The target aroma compounds of ALE in *Evolution niche E* (2+3-methylbutanol and isoamyl acetate) were more abundant in fermentations with the E2 strain than with the parental strain. 2+3-methylbutanol represents the combined signal for 2-methyl-1-butanol and 3-methyl-1-butanol that cannot be distinguished by this type of analysis. Phenylethyl alcohol and Phenylethyl acetate were more abundant in fermentations with strain G3 than with the parental strain. (B) PCA analysis of aromatic and branched amino acids' derived aroma compound profiles of natural wine must fermentations with parental and evolved strains. PC1 represents the variation between the parental and evolved strains and PC2 represents the variation of strains evolved in *Evolution niches G* (G3) and *E* (E2). The evolution points in opposite directions from the parental strain and generates a unique aroma profile.

Throughout the process of selecting isolated strains for phenotyping we were comparing the growth kinetics of the evolved isolates with their population of origin. However, the theoretical backbone of *EvolveX* depends on the fact that through the evolution process the fitness of the evolved strains will improve compared to the fitness of the starting parental strain. This was indirectly observed during the serial transfer experiment, since the initial inoculum was reduced by half and the time between the transfers was reduced between 25 and 50%. Nevertheless we chose to compare the difference in growth kinetics between 2 evolved strains from each evolution niche and the parental strain (Figure 3.9). Indeed, we observed that the parental strain exhibited impaired growth in both evolution niches. In more detail, in evolution niche E, the evolved strains were able to grow between to 3-4 times faster than the parental strain and the value of their maximum optical density was at least 3 times higher as well. The difference between the parental and the evolved strains was even more prominent in the evolution niche G, where the parental didn't exhibit considerable growth when the inoculation optical density was equal to OD_{600} 0.1. On the other hand, the two

evolved strains exhibited the phenotype based on which they had been initially selected for further characterisation.

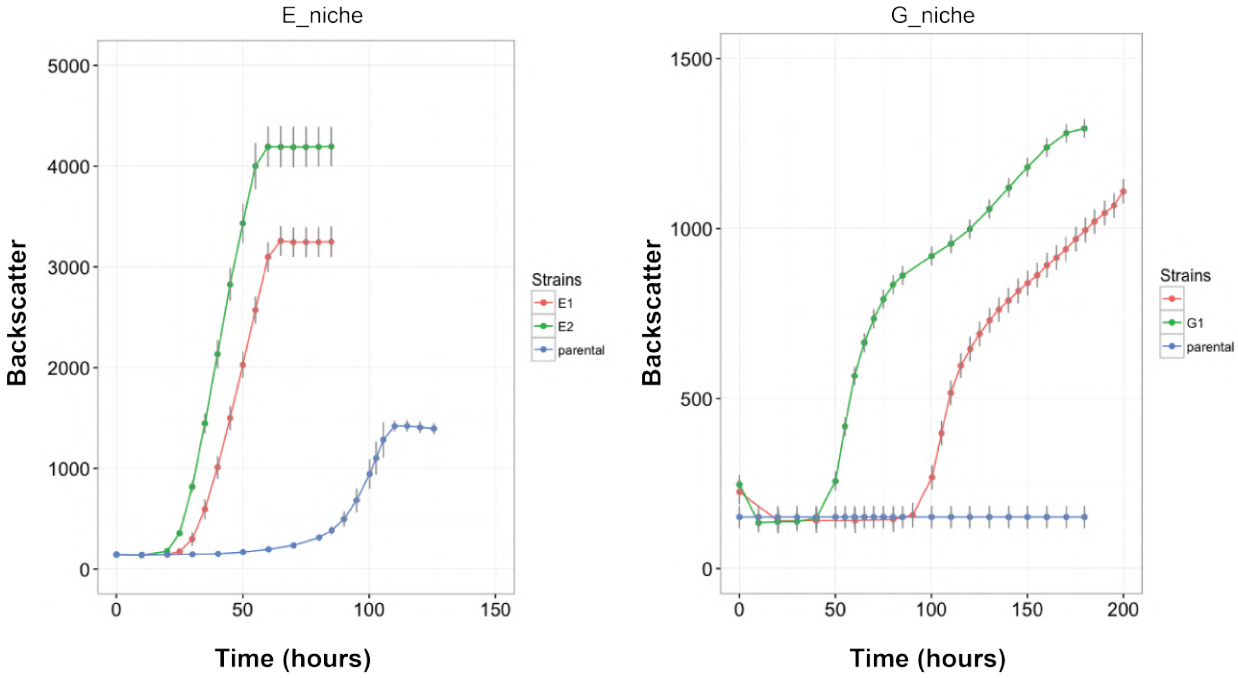

Figure 3.9. Comparison of the growth kinetics of the parental and evolved strains in the evolution niches. (A) Two evolved strains form the second round of evolution and their parental in the evolution niche E and (B) two evolved strains form the second round of evolution and their parental in the evolution niche G. All the evolved strains exhibit improved ability to efficiently grow in the evolution niches, while the parental strain exhibits impaired growth when the starting OD_{600} is equal to 0.1.

3.2.4.1 Summary

Our experiments so far showed that it is possible to perform ALE in an environmental niche designed by *EvolveX*. Specifically, we observed improvement of the overall fitness (based on measurements of the optical density) in all the lineages that we evolved in either one of the evolution niches. Moreover, despite the difficulties that we encountered due to the nature of our target compounds and the technical limitations concerning their detection, we were able to show that the evolved clones that were tested exhibited improved phenotype when they fermented NWM anaerobically. The phenotypic improvement was specific for each strain depending the niche in which it had evolved. These results were very encouraging concerning the principal of how *EvolveX* can be applied. However, *EvovleX* is meant to be used as a tool of directed evolution, where the phenotypic changes are based on initial predictions on specific fluxes and genes associated with them. For this reason, our next step was to try to identify whether these predictions came true during the evolution process.

3.2.5 Whole genome sequencing analysis

3.2.5.1 Analysis of samples originating from evolution niche E

Since the phenotypes of the evolved strains had changed in the desired way, we went on to sequence the genomes of evolved populations and isolated strains, to identify the changes that occurred at the genome level in comparison to the original parental strain. Through the analysis of the genomes of the populations and the strains that originated from evolution niche E, we were able to identify a few recurrent single nucleotide variations (SNVs) and specifically three different ones each in a different locus of gene *SKY1*. This gene is coding for a serine/threonine kinase, with threonine being one of the nitrogen sources in the evolution niche E. Also, the protein Sky1p interacts with Ptk2p, which is involved in spermine utilization (Figure 3.10A). Mutations in the gene *SKY1* were prevalent, with the allele frequencies reaching 0.97 in single strains and either 0.5 or 1 in population samples. Another interesting set of SNVs are two missense mutations in genes *LDB19* and *RSP5* which were occurring together. In more detail, the two proteins Ldb19 and Rsp5 interact, with Ldb19p being a regulator of endocytosis of plasma membrane proteins by recruiting the E3 ubiquitin ligase Rsp5p (Figure 3.10B). One of the target proteins for Rsp5p is Can1p, a known amino acid transporter (Lin *et al.*, 2008). In single strains the mutations in *LDB19* were homozygous, but the observed frequencies for *RSP5* were equal to 0.5, meaning that these strains are heterozygous for this variant. Apart from the described SNVs, the analysis also showed that both populations of the second lineage and single strains isolated from this population exhibited increased copy numbers of chromosome IV, indicating the emergence of aneuploidy for this chromosome (Figure 3.10C).

3.2.5.2 Analysis of samples originating from evolution niche G

In contrast, populations and isolated strains originating from evolution niche G, did not exhibit any recurrent SNVs, with the best example being the identification of two SNVs in the coding regions of genes *YKU70* and *REA1*. Even though these mutations were also detected in the population from where the strain was isolated, there is clear physiological role that can be associated between them and the targets of *EvolveX*. In more detail, Yku70p is involved in the maintenance of telomere length (Stellwagen *et al.*, 2003) and double-strand break repair, whereas Rea1p is involved in the biogenesis of ribosomes (Boulton and Jackson, 1996, Nissan *et al.*, 2004).

The most prevalent changes that were observed in these samples were associated with copy number variations (CNVs) (Figure 3.10C). In the third lineage evolved in niche G, as

well as in the strain isolated from this lineage the majority of the genome had triplicated, leading to the emergence of triploid strains. It is worth mentioning that this strain is the one that was phenotyped in NWM and exhibited the improved production of phenylethylacetate. The majority of this extensive increase in genetic content was detected in the first lineage, suggesting that adaptation that lead to aneuploidy were favoured under the conditions of evolution niche G.

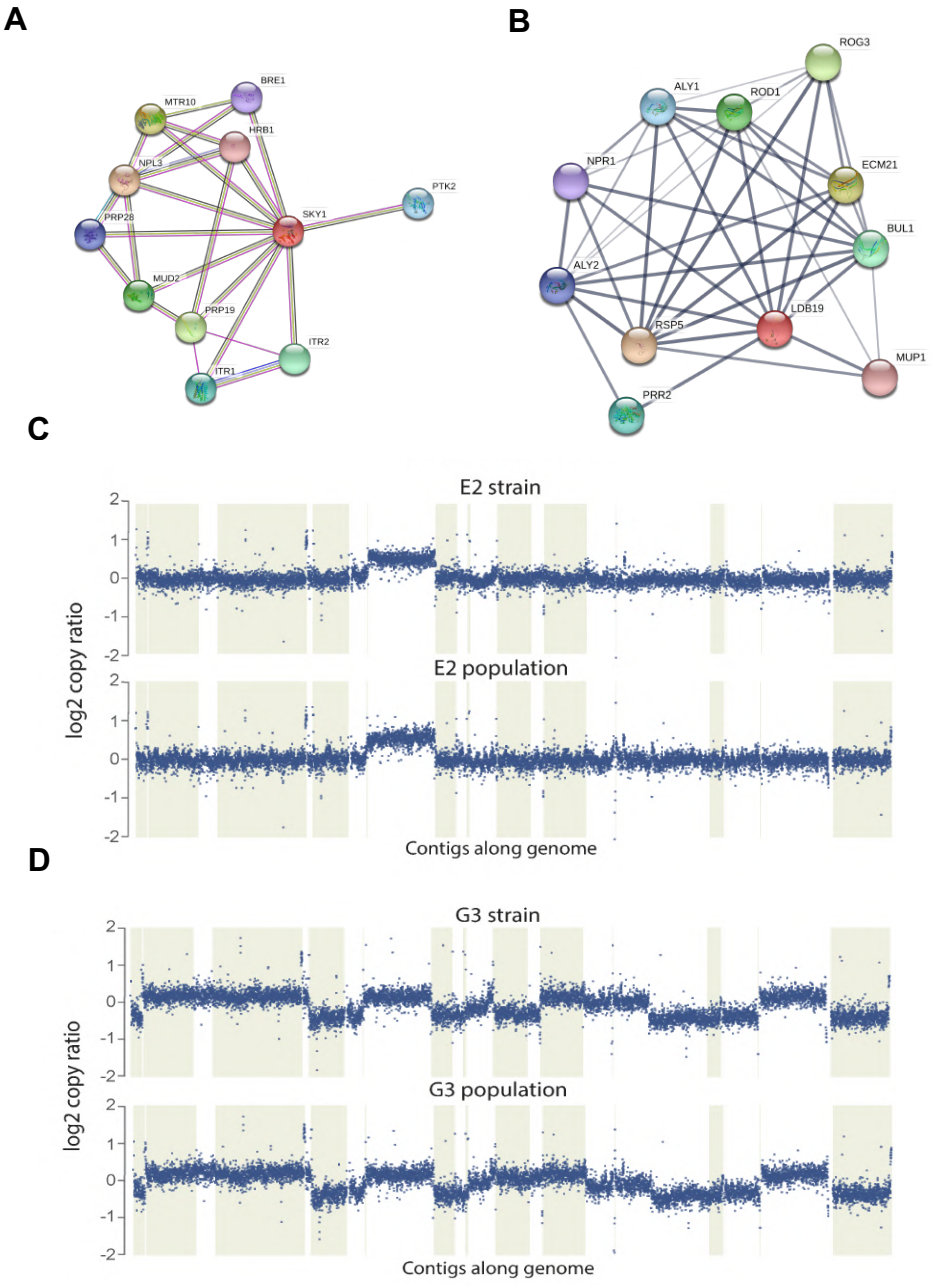

Figure 3.10. Evolved genome sequences harbour few SNVs and major CNVs. Protein-protein interaction networks (STRING database) of the proteins (A) Sky1p and (B) Rsp5p. Mutations of the two interacting proteins Rsp5p and Ldb19p were found together. (C) These chromosomes have enriched recurrent SNVs in E strains and populations. (D) G strains and populations encountered major chromosome copy number aberrations. Log2 copy ratios to parental strain (normalised to median) shown for G3 population and G3 strain isolated thereof.

3.2.6 Proteomics analysis in the evolution niches

Quantitative proteomics analysis was performed on evolved strains that were cultured in the evolution niches. Apart from the strains that we had already phenotyped, we selected one more strain from each evolution niche. All the strains we selected originated from lineages for which we had previously detected genomic evidence, aiming to better understand the mechanism responsible for the observed increased cell fitness.

In the case of Evolution niche E, only a small number of proteins was found to significantly differ (log_2folg change > 1, p-value < 0.05) between the parental and the evolved strains, but still PCA analysis of the proteome showed that the evolved strains are closer together than with the parental strain (Figure 3.11A). In more detail, a total of 24 (E1) and 16 (E2) proteins were over or under expressed in comparison to the parental strain (Figure 3.11B). The majority of these proteins (E1:13, E2:10) were identified as enzymes that participate in amino acid metabolism and transport. The biosynthetic pathways for L-arginine had decreased abundance, a fact that can easily be explained by the use of this amino acid as a nitrogen source in the medium. Increased utilisation of the available amino acids can be associated with increased abundance of the proteins Can1p and Put1p. Interestingly Can1p was more abundant only in the strain for which SNVs in genes *LDB19* and *RSP5* were detected. This result is intriguing, since it has been previously described that the protein Ldb19 recruits the ubiquitin ligase Rsp5 to Can1p to mediate its endocytosis (Lin *et al.*, 2008). Similarly, protein Sky1p was found to be less abundant in the strain that had SNV in the gene coding for this protein. Moreover, the use of ethanol as a carbon source explains the observed increase in the levels of alcohol dehydrogenase. The rest of the detected differences can be associated with nitrogen metabolism, decreased production of ammonia and export of the amount already present inside the cell.

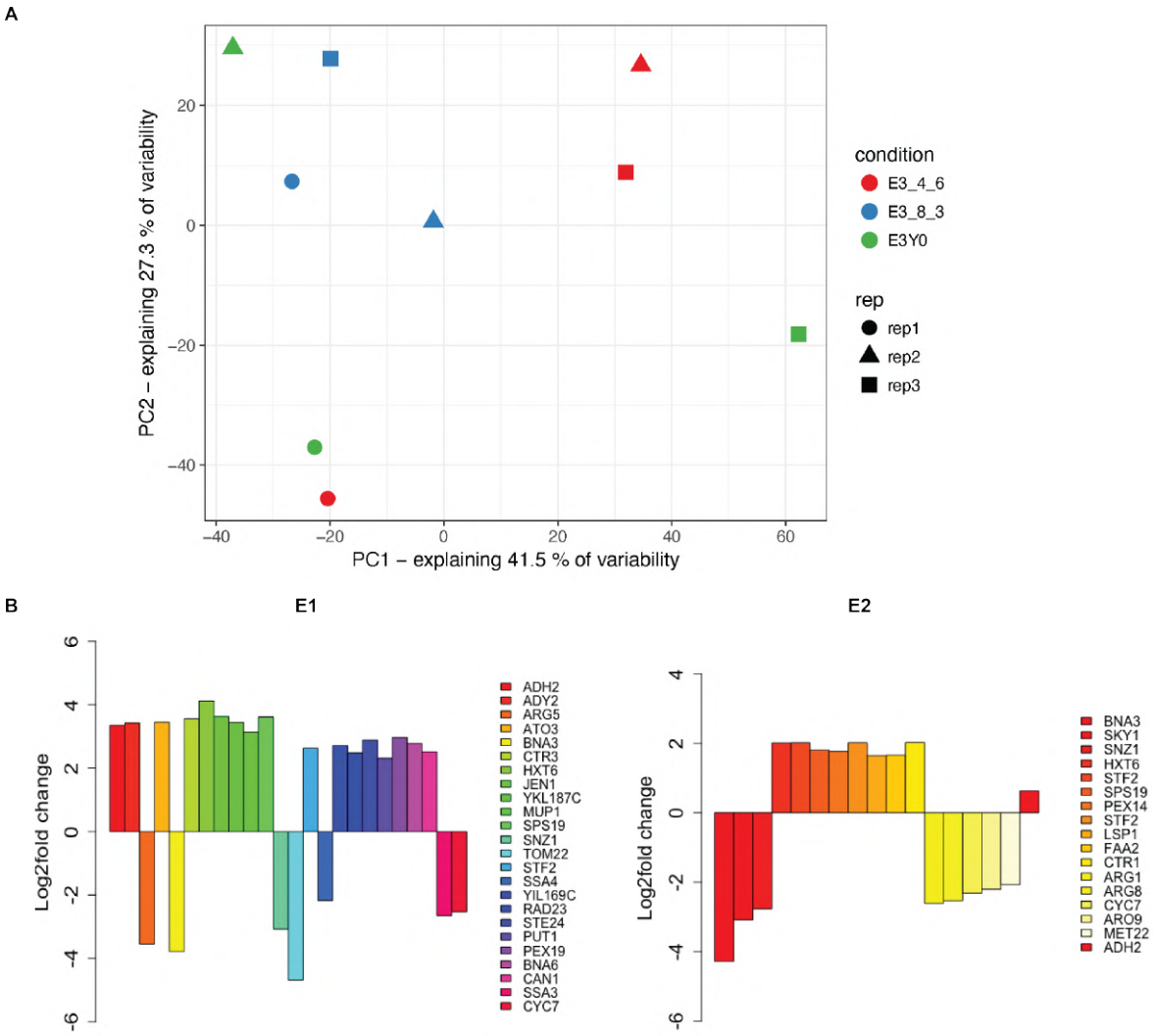

Figure 3.11. Proteomics analysis in the evolution niche E. (A) Principal component analysis of the whole proteome of the parental and the evolved strains. Most of the replicates originating from evolved strains are separated from their parental (E3Y0) in PC2, while the evolved strains (E3_4_6 and E3_8_3) are separated in PC1. (B) Differences in the expression (log_2fold change compared to the parental) of proteins associated with the evolution niche E, from strains E1 (E3_4_6) and E2 (E3_8_3).

In evolution niche G, the number of proteins, for which a change in abundance could be detected. was larger (G1: 114 and G3: 102) than what was already described for evolution niche E. This increased difference in the total proteome is depicted in a PCA plot (Figure 3.12A), where the evolved strains cluster together, separately from the parental strain. About one third of the significantly different proteins were metabolic enzymes, but most importantly enzymes involved in respiration were found to be enriched. The increase in respiration comes in agreement with the initial predictions of *EvolveX*, when the evolution niches were designed. Moreover, in strain G3, three glycolytic enzymes were lower in abundance, a fact that further supports the observation that these strains increased their respiratory activity (Figure 3.12B). Finally, in strain G1 higher abundances for proteins Stl1p and Aro8p were detected, with Stl1p being a glycerol proton symporter of the plasma

membrane and Aro8p being an aromatic amino acid transaminase, which is active in the presence of phenylalanine. The described differences are depicted on a metabolic map (Figure 3.12C), where it is shown that during the evolution process the strains had to fine tune their metabolism, as different parts of the same pathway were found to be either up- or downregulated.

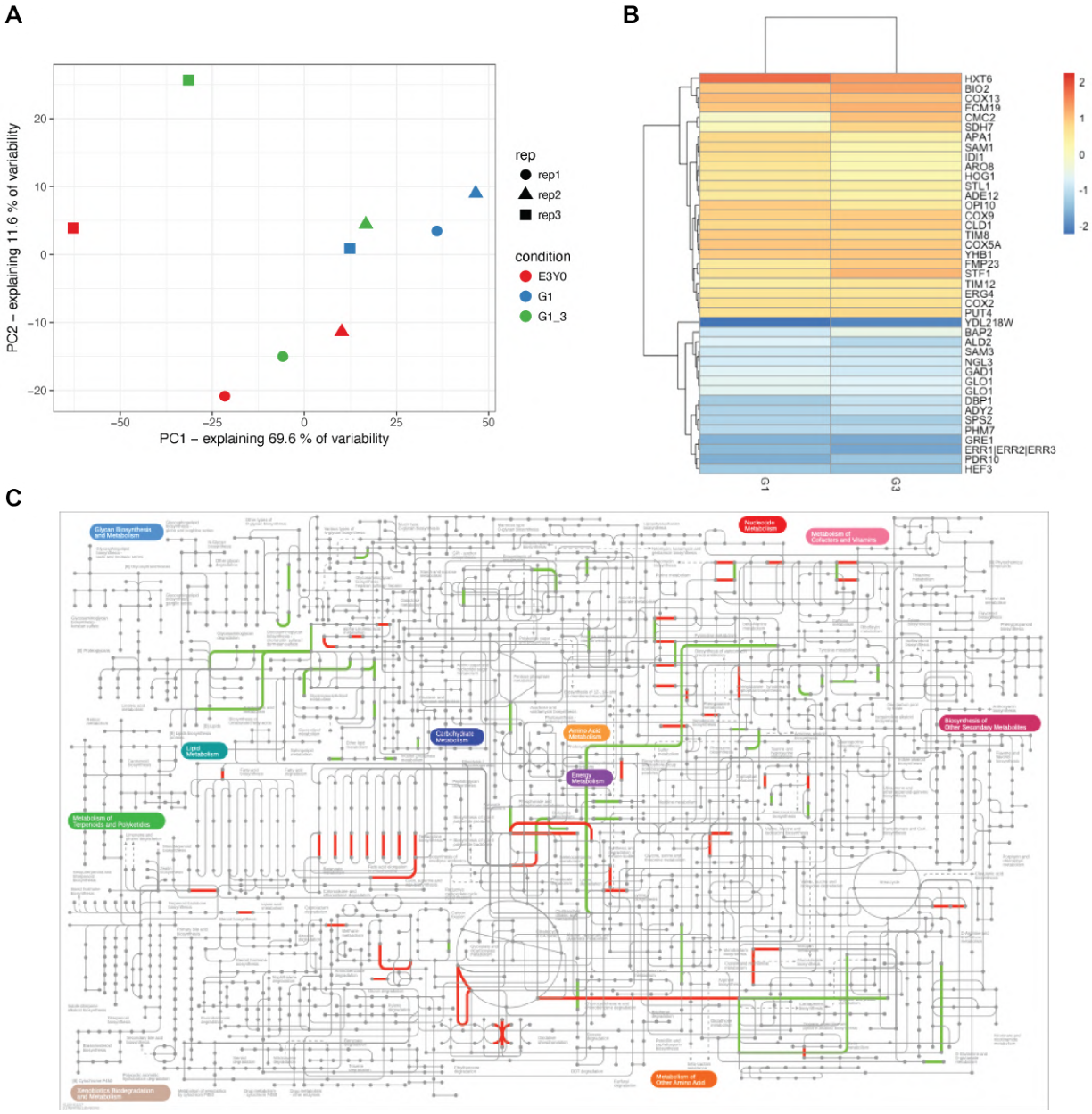

Figure 3.12. Proteomics analysis in the evolution niche E. (A) Principal component analysis of the whole proteome of the parental and the evolved strains. The evolved strains are separated from their parental in the first principal component. (B) Heatmap of the $\log_2$fold change between evolved strains and their parental strain for the proteins that can be associated with evolution niche G. The colour gradient is representing the value of the calculated $\log_2$fold change. (C) Mapping of the upregulated (red) and downregulated (green) proteins of the evolved strain G1 (previously phenotyped, Figure 3.8), on the metabolic map.

3.2.7 Transcriptomics analysis in the evolution niches

In opposite to the protein abundances, the number of genes with significantly changed expression levels is higher in the strains from Evolution niche E than G. Strains from the Evolution niche E differentially expressed on average more than 200 genes (E1: 203 and E2: 235), while stains from the evolution niche G differentially expressed less than 100 genes (G1: 72 and G3: 49). A limited overlap in differentiated gene expression between the strains was observed (E1: 13, E2: 7, G1: 6, G3: 4), which can be associated with the different conditions in which the strains were grown. PCA analysis of the evolved strains and their parental showed that in both evolution niches the evolved strains cluster away from the parental strain. In both cases, the first principal component explains 63% of the variability and separates the parental and the evolved strains, while the difference between the evolved strains in each case is explained by the second principal component (Figure 3.13).

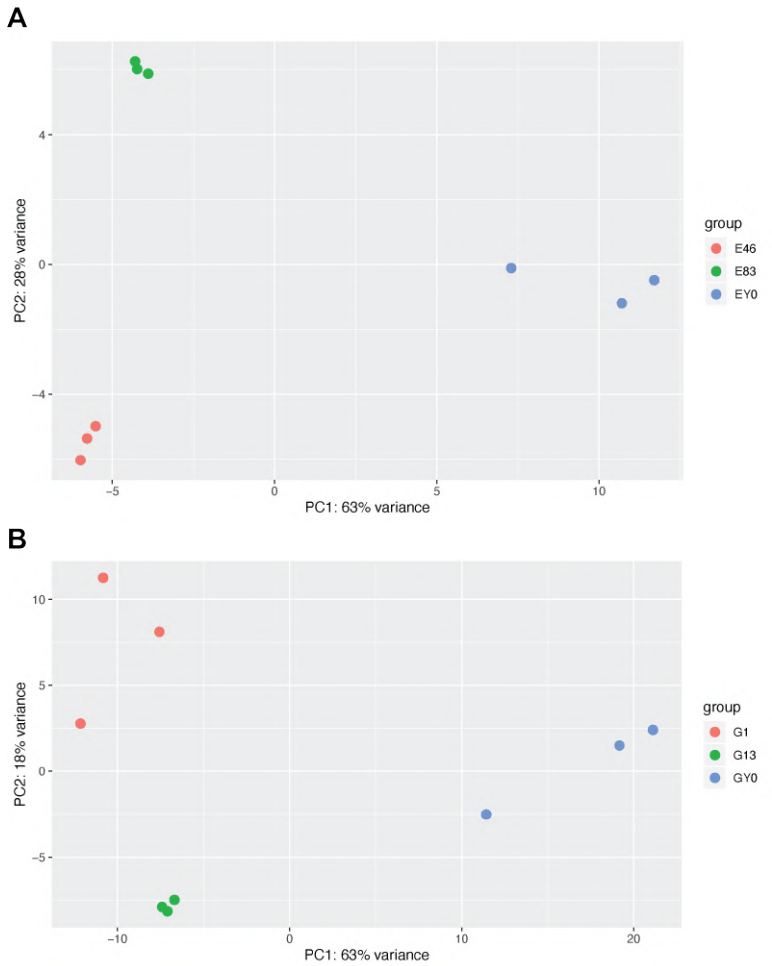

Figure 3.13. Principal component analysis of the parental and the evolved strains, based on their transcriptome in the evolution niches. (A) In evolution niche E the evolved strains are separated from the parental strain from the first principal component (63% of the variance), while the second principal component (28%) separates the two evolved strains. (B) In evolution niche G the evolved strains are separated from the parental strain from the first principal component (63% of the variance), while the second principal component (18%) separates the two evolved strains.

3.2.8 Evaluation of the transcriptome and the proteome in the target niche

One of the main hypotheses of *EvolveX* was that the differences in phenotype that would occur during the evolution process would remain when the evolved strains would be reintroduced into the target niche. We gathered strong evidence for this through the metabolomics analysis of NWM fermentations, however, we also wanted to understand the mechanistic differences that occur and lead to the increased aroma production during the fermentation process. For this purpose, we conducted RNA sequencing and quantitative proteomics experiments in small scale NWM fermentations. The test fermentations that we performed showed that the proper collection time was between 2.5 and 3 days (Figure 3.14).

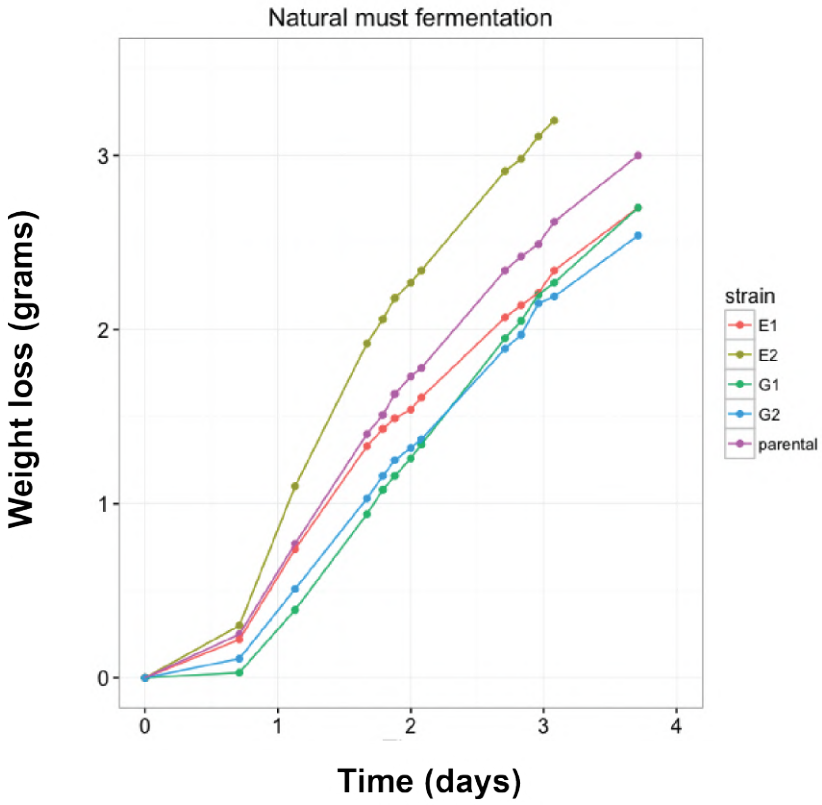

Figure 3.14. Growth kinetics during the fermentation of NWM. The parental and four evolved strains were grown anaerobically and their growth was estimated as weight loss over time, which represents the amount of CO_2 released in the environment. In the figure the average value of biological replicates is represented

The number of proteins that differed significantly between the evolved strains and the original parental strain was quite smaller compared to the number of proteins that were identified in the evolution niches (E1: 3, E2: 7, G1: 8, G3: 7), but even so, some of them could be used to draw conclusions. Increased abundance of protein Tkl1p in strain G3 can be associated with the increased production of phenylethyl acetate. Tkl1p is a transketolase, which is needed for the synthesis of aromatic amino acids, as it catalyses the conversion of

xylulose-5-phosphate and ribose-5-phosphate to sedoheptulose-7-phosphate and glyceraldehyde-3-phospate in the pentose phosphate pathway. Another interesting observation is that the abundance of protein His1p was decreased, since this protein is involved in the biosynthesis of the amino acid histidine and this pathway competes with the pathway of aromatic amino acids synthesis for the essential precursor ribose 5-phosphate. Similar changes in the abundance of these two proteins were observed in strain G1 as well (Figure 3.15A).

These observations are supported by changes in gene expression ((Figure 3.15C), with most noticeable the changes in expression of genes being associated to the synthesis of chorismate (*ARO1*, *ARO3*, *RK11*), the biosynthesis of aromatic amino acids (*ARO8*) and the decarboxylation of phenylpytuvate (*PDC1*, *PDC5*). Decreased levels of gene expression had occurred in pathways linked to the synthesis of histidine and purines, which as described above for the proteomics analysis, require the same precursors as the aromatic amino acid synthesis pathway. All the above-mentioned examples are among the original *EvolveX* predictions that were used for the creation of the evolution niche G.

The strains evolved in niche E showed increased abundance for the protein Tkl1p, though without observed differences in the levels of His1p (Figure 3.15A). This observation may potentially be due to the fact that the reaction catalysed by this enzyme leads both to increased levels of phenylethanol and the pool of aroma compounds such as 2-methyl-1-butanol and 3-methyl-1-butanol which derive from the branched-chain amino acids (Figure 3.15B). An important finding is that strain E2, which was the strain that we phenotyped, has increased levels of the protein Ilv3p, a dihydroxy acid dehydratase, which catalyses the third step in the biosynthetic pathway of branched-chain amino acids and of Leu1p which is involved in the biosynthetic pathway of leucine. Abundance differences for both of these proteins agree with the initial predictions of *EvolveX*. A final association between the initial computational predictions and the experimental evidence is the protein Leu2p, which was predicted to be upregulated in the case of evolution niche E, but downregulated in the case of evolution niche G. Indeed, the levels of Leu2p were decreased in the samples from strain G1, but remained unchanged in the samples of strains E1 and E2.

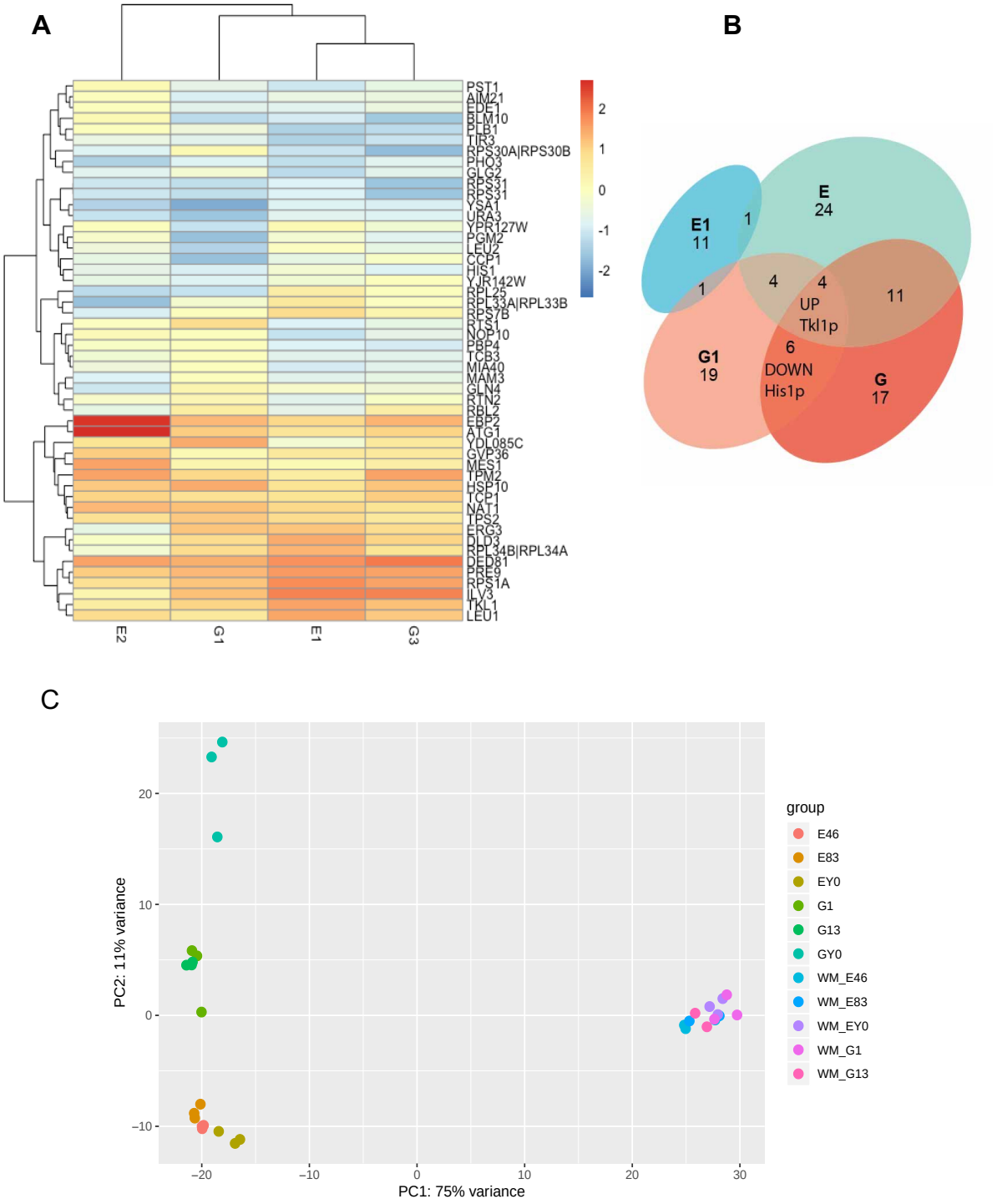

Figure 3.15. Proteomics and transcriptomics of the parental and the evolved strains in the target niche. (A) Heatmap of the log_2fold change between evolved strains and their parental, based on their proteome when cells are cultured in NWM. Two of the evolved strains originate from the evolution experiment in the evolution niche G and the other two from the evolution experiment in the evolution niche E. The colour gradient is representing the value of the calculated log_2fold change. (B) Euler diagram of differentially abundant protein sets (from the analysis in NWM) in the evolved strains E1 and E2, from *Evolution niche E*, and the evolved strains G1 and G3, from *Evolution niche G*. (C) Principal component analysis of the parental and the evolved strains, based on their transcriptome in NWM and comparison to the transcriptomics analysis in the samples from the evolution niches. All the samples originating from NWM cluster together, independently of their origin (parental or evolved) and are separated from the samples from the evolution niches (first principal component, 75% of the variance).

3.2.9 Summary and outlook

Natural selection shapes the characteristics of a population by acting on the genotypes of individuals increasing the ability of the best adapted ones to survive and proliferate. During such events, specific genotypes can exapt for novel traits, which offer an adaptive advantage under certain environmental conditions. In the present study, we developed the algorithm *EvolveX* which uses genome-scale metabolic model simulations to identify which chemical environment composition (evolution niche) would create a relative selection pressure on specific metabolic fluxes, towards the production of a target compound. Following, the computed chemical composition will be used as the environment for cellular adaptation with the use of ALE. The adaptation of the metabolic network to such an evolution niche will result into increased production of the target compound in the natural environment (target niche). This approach can solve the problem of using laboratory evolution to acquire traits that do not increase cellular fitness directly. We successfully implemented *EvolveX* to increase the production of organic compounds associated with the aroma profile of wine, in wine strains of *S. cerevisiae*. Moreover, we showed that the improvement of the phenotype towards a specific target is specific, since strains that evolved in the Evolution niche E were not producing higher amounts of phenylethyl acetate, which was the target compound of evolution niche G. Moreover, the evolution niches direct specific parts of the metabolism towards the desired outcome without largely affecting the rest, as we observed both from the strains' ability to grow normally in WMM and from the fermentation profile comparison between the parental and the evolved strains. This observation is important, because the use of *EvolveX* allowed the improvement of specific phenotypic traits, while the rest of the important attributes remained intact and thus the evolved strains can readily be used in large-scale industrial settings. Apart from increased production of aroma compounds in NWM, the evolved strains exhibited also increased fitness in the evolution niches compared to their parental strain. Thus, we proved that the production of aromatic compounds was linked to improved growth in the selected evolution niches, which was one of the initial assumptions of *EvolveX*.

Even though we were not able to pinpoint specific causative mutations with use of whole genome analysis, analysis of the proteome and the transcriptome proved useful to understand the changes that occurred during the evolution process. Moreover, these kinds of analyses provided evidence about the success of *EvolveX* to accurately predict the changes that would occur during adaptation in the evolution niches. For example, the samples originating from the evolved strains from niche G were enriched in enzymes involved in

respiration, which had been predicted by simulating the selection pressure on each metabolic reaction individually in the particular Evolution niche. Interestingly, strains that were evolved in the same niche exhibited different up and downregulated protein profiles in the same pathways. This phenomenon was observed in strains that evolved in niche E as well, as there identified different proteins of the Arginine biosynthesis pathway that were downregulated (E1: Arg1p, Arg8p and E2: Arg5p). Moreover, different subunits of cytochrome c oxidase were upregulated in the G1 and G3 strains. This evidence suggests that during the evolution process different cell lineages adopt different adaptation strategies which lead to the same endpoint (phenotypic change) and thus accurate prediction of each single adaptation is very challenging. In the samples originating from NWM fermentations we were able to identify gene expression changes that supported the initial predictions of *EvolveX* (Figure 3.16), such as the differently regulated gene *TKL1* which was a target in both evolutionary niches, or the protein Ilv3p which is involved in the biosynthesis of branched-chain amino acids and so it is required for the production of higher alcohols (evolution niche E). In conclusion, *EvolveX* is the first model guided method for evolving non-growth associated traits, which can be generally applied for the improvement of strains and offer solutions for non-GMO applications.

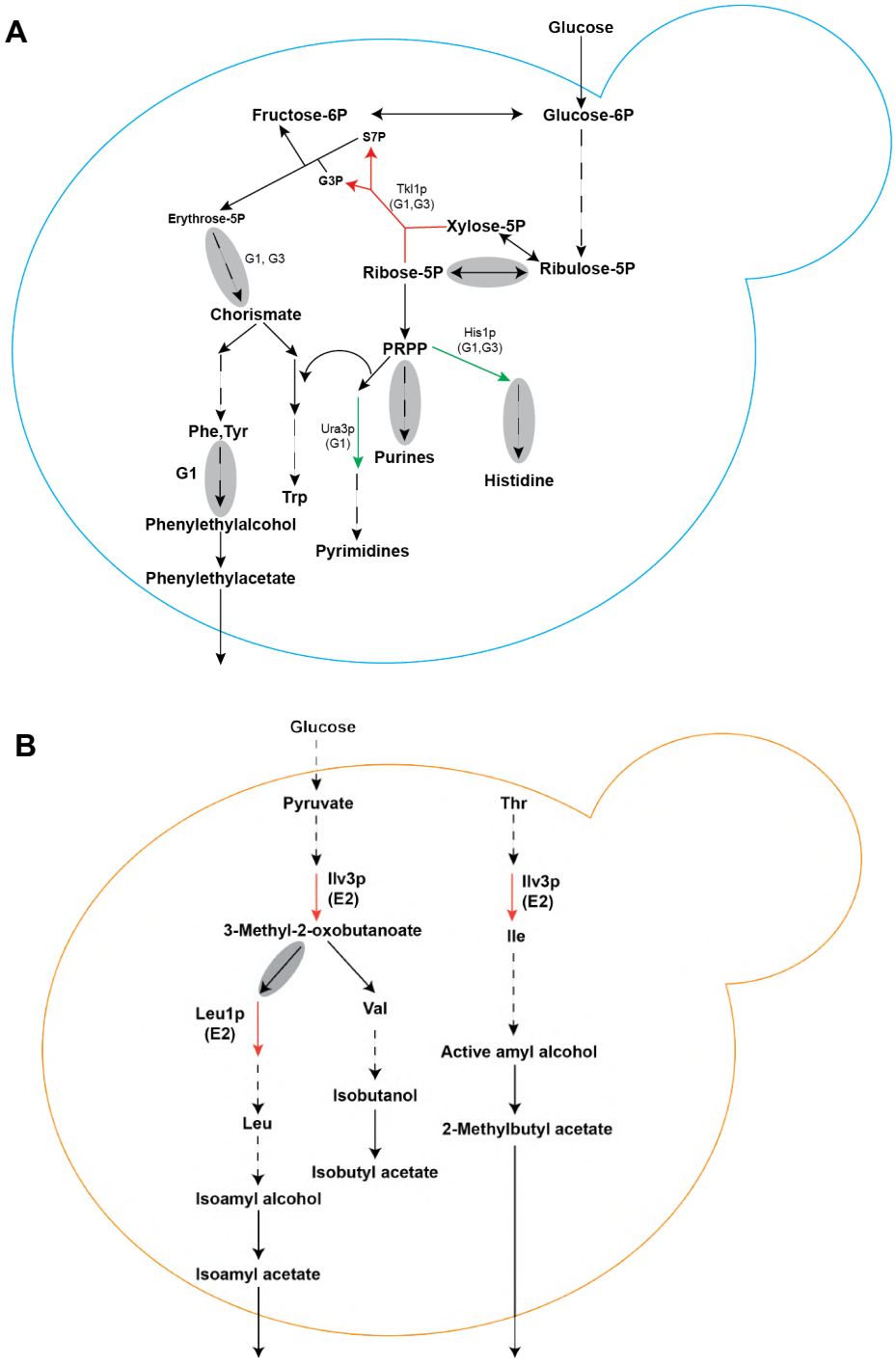

Figure 3.16. Schematic representation of the changes that occurred during the evolution process. (A) Evolved alterations in metabolic pathways towards phenylethyl alcohol and phenylethyl acetate generation in natural wine must fermentations. The differentially abundant proteins are indicated with coloured arrows, (red for upregulation and green for downregulation) whereas the gene expression level changes are indicated with grey clouds. Dashed lines represent pathways of several reactions. (B) Evolved alterations in metabolic pathways towards 2-methyl-1-butanol (i.e. active amyl alcohol) and 3-methyl-1-butanol (i.e. Amyl alcohol) and their acetate esters' (i.e. isoamyl acetate and 2-methylbutyl acetate) generation in natural wine must fermentations.

3.3 Material and methods

3.3.1 Model simulations •

S. cerevisiae consensus genome-scale metabolic model v. 7.6 (Lu *et al.*, 2019) was used with few revisions. For implementing the model revisions, Matlab R2017b v. 9.3.0 and Cobra toolbox v.3.0 (cloned 29.03.2018) (Heirendt *et al*. 2019) was used. *EvolveX* algorithm was implemented and excecuted in Matlab R2015a v. 8.5.0 with IBM ILOG CPLEX v. 12.6.1 functions 'cplexlp' and 'cplexmilp"

3.3.2 Strains and culture media

The parental strain used for ALE was the commercially available diploid wine strain *S. cerevisiae* QA23, which was obtained from Lallemand.

Adaptive laboratory evolution was performed in two separate evolution niches. The two media consist of the same mineral concentration and different carbon and nitrogen sources. Both media contained 6.6 g/L K_2SO_4, 3 g/L KH_2PO4, 0.5 g/L $(MgSO_4)7H_2O$, 1 mL/L vitamin solution, 1 mL/L trace minerals solution and the pH was set to a value of 6. The vitamin solution is composed of 50 mg/L d-biotin, 200 mg/L para-amino benzoic acid, 1.0 g/L nicotinic acid, 1.0 g/L Ca-pantothenate, 1.0 g/L pyridoxine-HCl, 1.0 g/L thiamine-HCl and 25 mg/L *myo*-inositol and the trace minerals solution of 3 g/L $FeSO_4 \cdot 7H_2O$, 4.5 g/L $ZnSO_4 \cdot 7H_2O$, 4.5 g/L $CaCl_2 \cdot 6H_2O$, 0.84 g/L $MnCl_2 \cdot 2H_2O$, 0.3 g/L $CoCl_2 \cdot 6H_2O$, 0.3 g/L$CuSO_4 \cdot 5H_2O$, 0.4 g/L $NaMoO_4 \cdot 2H_2O$, 1 g/L H_3BO_3, 0.1 g/L KI and 15 g/L$Na_2EDTA \cdot 2H_2O$.

The first of the media which was named Evolution niche E contained 7.5 g/L Ethanol as a carbon source, 1.7 g/L Arginine and 0.8 g/L Glycine as nitrogen sources. The second medium named Evolution niche G contained 5 g/L glycerol as carbon source, 5g/L Phenylalanine and 1.2 g/L Threonine as nitrogen sources. The media were sterilized through filtration with use of 0.22 μm filters.

Prior to genomic DNA extraction for whole genome sequencing or inoculation in the evolution niches the selected strains were grown overnight in rich medium (YPD) containing 10 g/L of yeast extract, 20 g/L peptone and 20 g/L glucose which was sterilized through autoclaving. Single colonies for the start of the experiments were either picked from YPD or wine must mimicking medium (WMM) plates that contained 2% [w/v] agar as solidifying agent.

3.3.3 Adaptive laboratory evolution

At the start of the adaptive evolution experiment, 6 single colonies of the parental QA23 strain were inoculated in rich medium overnight and these 3 of them were inoculated in each of the Evolution niches media with a starting OD_{600} equal to 0.2. The serial transfer experiment was performed with 7 mL liquid cultures in 50 ml shake flasks in triplicate lineages on each medium. The Erlenmeyer flasks were capped with cotton plugs to enhance aeration. Cultures were incubated at 30 ^{o}C with shaking at 180 rpm. When an increase in OD was visually observed, the OD_{600} of each flask was measured in a spectrophotometer and the value was used to calculate the appropriate volume to be transferred into fresh medium with a starting OD_{600} of 0.2. When the populations began to exhibit improved growth rates and higher OD_{600} the volume transferred was calculated to be appropriate for a starting OD_{600} of 0.1. Population stocks were collected after every transfer and an appropriate volume of each replicate was mixed with 80% [w/v] glycerol to a final volume of 30% [w/v] and stored at -80 ^{o}C. The number of generations was calculated back using the following formula $[Log_{10}(A_f/A_i)]/0.3$, where A_f is the OD_{600} before the transfer and A_i is the OD_{600} that was initially inoculated.

3.3.4 Characterization of evolved strains

The end point populations of all the lineages were platted on WMM + 2% [w/v] agar plates and single colonies were isolated after 48 h. Specifically, nine single colonies were isolated from each final population and they were grown overnight in liquid cultures of WMM. These overnight cultures were used to prepare glycerol stocks of each isolate and were inoculated in the appropriate Evolution niche medium with a starting OD_{600} of 0.1. The growth of the isolates was compared to the growth of the end point population on the evolution medium in the same conditions as in the ALE experiment, 30 ^{o}C with shaking at 180 rpm. Cell growth was monitored with OD_{600} measurements at regular time intervals. The best performing strain from each lineage was chosen for further characterisation in wine fermentation mimicking conditions.

Fermentations and analysis of the volatile compounds were performed for the parental and isolated evolved strains from both Evolution niches. The experiments were performed at ICVV (Instituto de Ciencias de la Vid y del Vino), Logrono, Spain, by Dr. Ramon Gonzalez, according to Curiel *et al.* (2016).

3.3.5 Whole genome sequencing of populations and isolates

Genomic DNA was extracted from parental strains grown in YPD, lineages grown in Evolution niches, and single isolated strains grown in YPD using phenol-chloroform based extraction. Specifically, a total volume of 7 mL overnight cultures was centrifuged at 3000 rpm for 3 min and the pellets were washed with sterile ddH2O. The cells were resuspended in 2 mL of TrisEDTA solution (0.1 M Tris and 0.1 M EDTA) and transferred to Eppendorf tubes, with the addition of 15 μL of lyticase. The pellets were then incubated at 37 ^{o}C for 30 min. Next the spheroplasted cells were centrifuged at 1500 rpm for 2 minutes (Eppendorf centrifuge), the supernatant was removed and the cells were resuspended in 400 μL of breaking buffer which contained 10mM Tris, 1mM EDTA, 100mM NaCl, 2% [v/v] Triton X-100 and 1% [w/v] SDS. The cell suspensions were transferred to FastPrep Cap tubes tubes with 200 μL of glass beads (400 nm acid washed, Sigma) and the cells were lysed with 3 rounds of bead beating at 4.5 MHz/sec for 20 seconds with 1 min cooling intervals. The cell lysates were transferred to a new tube that contained 400 μL of phenol-chloroform/isoamylic alchocol and 400 μL of TE buffer (Tris 50 mM, EDTA 20 mM) and were centrifuged briefly until an emulsion was formed. The emulsions were centrifuged at 13.000 rpm for 5 minutes at RT. The aqueous phase of each tube was transferred to a new Eppendorf, it was mixed with 1 mL of cold 100% ethanol and incubated at room temperature for 10 min to help precipitation. In the next step, the tubes were centrifuged at 13.000 rpm for 5 minutes at room temperature, the ethanol was removed and the DNA pellet was resuspended at 400 μL of TE buffer with 2 μL of RNAse solution (20 mg/mL) and incubated for 15 minutes at 37^{o}C, followed by a second incubation step at 65^{o}C for 15 minutes, in order to deactivate the RNAse. The DNA solution was mixed with 400 μL of phenol-chloroform/isoamylic alchocol and the extraction step was performed again as described above. DNA was precipitated from the aqueous phase with 1 mL of cold 100% ethanol and centrifugation at 13,000 rpm for 5 minutes at room temperature. The pellet was left to dry for 30 minutes at 55 ^{o}C, next was resuspended with 50 μL of H2O and was left overnight at 4 ^{o}C for the pellet to dissolve completely. The quality of the extracted DNA was evaluated with electrophoresis in a 1% [w/v] agarose gel. DNA concentrations were measured using a Qubit (Thermo Fisher Scientific, USA). Equal amounts of DNA from all samples were used for library preparation, which was done with the NEBNext DNA Ultra2 Library Preparation Kit (New England Biolabs). The preparation of the library was performed on an automated liquid handling system (Hamilton Robotics), the quality of the library was tested on a 2100 BioAnalyzer (Agilent Technologies), and the DNA concentration was measured using a Qubit.

Sequencing was performed at the Genomics Core Facility (EMBL Heidelberg) with use of the HiSeq2500 platform (Illumina, San Diego, USA) and the run produced 250 bp paired-end reads.

3.3.6 Whole genome sequence data analysis •

The quality of the obtained reads was checked using Fastqc v. 0.11.4 (Andrews *et al*, 2010). Adapter removal and low quality read filtering was performed using cutadapt v. 1.9.1 (Martin et al, 2011). The trimmed reads were aligned to Saccharomyces cerevisiae EC1118 reference genome (Novo et al. 2009) with the Burrows-Wheeler Aligner v. 0.7.12 mem (Li and Durbin, 2009) using default parameters. The alignments were processed (added read groups, sorted, reordered, and indexed) and duplicate reads were marked using Picard Tools v. 1.129 (https://broadinstitute.github.io/picard/). Single nucleotide variant (SNV), and insertion-deletion (indel) variant calling was performed against the parental sample with GATK4 v. 4.1.0.0 Mutect2 using the S. cerevisiae EC1118 as the reference and default parameters. Panel of Normals for the variant calling was compiled of 47 wild type S. cerevisiae strains (winery isolates and commercial wine strains, including the parental) sequenced on the same platforms as the actual samples. Variant calling was first performed for the wild type strains by running Mutect2 in tumour-only -mode and then the panel of normal was created with GATK4 v. 4.1.0.0 CreateSomaticPanelOfNormals. The variant calls were filtered using GATK4 v. 4.1.0.0 FilterMutectCalls using default thresholds.

Copy number variant (CNV) analysis was performed on read counts and minor allele frequencies using GATK4 v. 4.1.0.0 tools. First the read counts were binned into 1000 bp intervals with CollectReadCounts. These read counts were denoised with DenoiseReadCounts using a panel of normal created from equally binned read counts of 47 wild-type strains with CreateReadCountPanelOfNormals. The allelic counts were collected using CollectAllelicCounts and combined with the binned read counts for modelling the CNV segments using ModelSegments with number-of-changepoints-penalty-factor of 4. CNVs were called using CallCopyRatioSegments. In major, directionally imbalanced, copy number aberrations the default centralization of the log2 copy ratios to median across all contigs, misplaces the zero level to deviate from the conserved copy number. As there were major differences in the copy number aberrations between the samples, the copy ratios were re-normalized to the level of segments with conserved copy number. These segments were identified from the minor allele frequencies and copy ratio differences between modelled segments.

3.3.7 RNA-sequencing sample preparation and data analysis •

All the RNA samples were prepared according to the following procedure. Total volume of 20 mL from each culture was transferred to a 50 mL Falcon® filled with ice and was immediately centrifuged at 3000 rpm for 3 min at 4 ^{o}C (Eppendorf centrifuge). Next the supernatant was discarded and the cell pellet was snap frozen into liquid nitrogen and stored at -80 ^{o}C, until the extraction was performed. Total RNA from the pellets was extracted with the RNAeasy kit (Qiagen) according to the manufacturer's recommendations. In brief, 594 µL of RTL buffer plus 6 µL of β-mercaptoethanol were used to resuspend the frozen cell pellet which was left on ice. The resuspended cells were transferred to an ice cold FastPrep Cap tube which contained 600 µL glass beads (400 nm acid washed, Sigma). The cells were then lysed with 2 cycles of bead beating, each cycle lasted 10 sec at 6 Mz/sec with 15 sec cooling interval. Cell lysates were transferred to a new tube and were centrifuged for 2 min at full speed (Eppendorf centrifuge) and the supernatant was carefully mixed with 1 volume of 70% [v/v] HPLC-grade ethanol. Next, the total volume of the sample was transferred to an RNAeasy column and the manufacturer's instructions were followed. Total RNA was eluted with 60 µL of RNAse free water and Turbo DNAse (Invitrogen Ambion) was used to digest leftover DNA according to the manufacture instruction. Finally, one more step of RNA clean-up was performed with the same kit.

The RNA library was prepared using the NEBNext® Ultra™ II Directional RNA Library Preparation Kit from Illumina: polyA transcripts capture. Briefly, barcoded stranded mRNA-seq libraries were prepared from high quality total RNA samples (~200 ng/sample) using the NEBNext Poly(A) mRNA Magnetic Isolation Module and NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs (NEB), Ipswich, MA, USA) implemented on the liquid handling robot Beckman i7. The obtained libraries that passed the QC step were pooled in equimolar amounts; 2 pM solution of this pool was loaded on the Illumina sequencer NextSeq 500 and sequenced uni-directionally, generating ~500 million reads, each 75 bases long.

The quality of the obtained RNA sequencing reads was assessed and summarized with Fastqc v. 0.11.5 (Andrews *et al*, 2010). Adapter trimming, to remove the standard lllumina TrueSeq Index adapters sequences, was performed using cutadapt v. 2.3 (Martin et al, 2011). Subsequently, quality read filtering and trimming was performed by FaQCs v. 2.08 (Lo and Chain, 2014), with the following parameters: `-q 20 -min_L 30 -n 3`. After the trimming and filtering steps, the total number of reads was, on average, 31 million.

• The following method parts were written for Jouhten[*], Konstantinidis[*] *et al.*, manuscript in preparation.

Trimmed reads were then aligned to the reference genome of *S. cerevisiae* EC1118 (EnsemblFungi: annotation number GCA_000218975.1) using STAR v. 2.5.2a (Dobin et al, 2013). On average, 85% of reads uniquely mapped to an annotated feature in the reference. Only uniquely mapped reads were then used to generate the gene level count tables with HTSeq v. 0.9.1 (Anders et al., 2015). Differential expression analysis, including multiple testing correction and independent filtering, was performed with Bioconductor package: DESeq2 v. 1.12.0 (Love et al., 2014). False discovery rate (FDR) was calculated with fdrtool v. 1.2.15 (Strimmer, 2008) using the raw p-values returned by DESeq2. Genes with a FDR < 0.05 were considered as significantly differentially expressed. Unless specified, all packages were used with default parameters. Biostatistical analysis was conducted with R v. 3.6.1 (R Development Core Team).

3.3.8 Protein extraction

For the extraction of total proteome 10 mL of each culture were transferred into ice-cold 15 mL Falcon® tubes which were centrifuged immediately at 3000 rpm for 3 min at 4 ^{o}C. The supernatant from the centrifugation was discarded and the cell pellets were washed once with 1 mL of cold PBS. The washed pellets were snapped frozen with liquid nitrogen and stored at -80 ^{o}C. For the extraction, the cell pellets were lysed with 0.1% [v/v] RapiGest (Waters) in 100 mM ammonium bicarbonate and mechanical disruption with 3 rounds of sonication of 10 sec sonication and 10 sec rest on ice per round. Sonication was followed by 2 cycles of bead beating (200 µL glass beads, 400 nm acid washed, Sigma), each cycle lasting 20 sec at 4 Mz/sec with 1 min cooling intervals between the cycles.

3.3.9 Sample preparation and TMT labelling *

Reduction of disulphide bridges in cysteine containing proteins was performed with dithiothreitol (56 °C, 30 min, 10 mM in 50 mM HEPES, pH 8.5). Reduced cysteines were alkylated with 2-chloroacetamide (room temperature, in the dark, 30 min, 20 mM in 50 mM HEPES, pH 8.5). Samples were prepared using the SP3 protocol (Hughes *et al.*, 2018) and trypsin (sequencing grade, Promega) was added in an enzyme to protein ratio 1:50 for overnight digestion at 37 °C. Peptides were labelled TMT10plex (Werner *et al.*, 2014). Isobaric Label Reagent (ThermoFisher) according the manufacturer's instructions. For further sample clean up an OASIS® HLB µElution Plate (Waters) was used. Offline high pH reverse phase fractionation was carried out on an Agilent 1200 Infinity high-performance liquid chromatography system, equipped with a Gemini C18 column (3 µm, 110 Å, 100 x 1.0 mm, Phenomenex), (Reichel *et al.*, 2016), resulting in 12 fractions.

3.3.10 Mass spectrometry data acquisition *

An UltiMate 3000 RSLC nano LC system (Dionex) fitted with a trapping cartridge (µ-Precolumn C18 PepMap 100, 5µm, 300 µm i.d. x 5 mm, 100 Å) and an analytical column (nanoEase™ M/Z HSS T3 column 75 µm x 250 mm C18, 1.8 µm, 100 Å, Waters). Trapping was carried out with a constant flow of trapping solution (0.05% [v/v] trifluoroacetic acid in water) at 30 µL/min onto the trapping column for 6 minutes. Subsequently, peptides were eluted via the analytical column running solvent A (0.1% formic acid in water) with a constant flow of 0.3 µL/min, with increasing percentage of solvent B (0.1% [v/v] formic acid in acetonitrile) from 2% to 4% in 4 min, from 4% to 8% in 2 min, then 8% to 28% for a further 37 min, in another 9 min. from 28%-40%, and finally 40%-80% for 3 min followed by re-equilibration back to 2% B in 5 min. The outlet of the analytical column was coupled directly to an Orbitrap QExactive™ plus Mass Spectrometer (Thermo) using the Nanospray Flex™ ion source in positive ion mode.

3.3.11 MS method for TMT10 *

The peptides were introduced into the QExactive plus via a Pico-Tip Emitter 360 µm OD x 20 µm ID; 10 µm tip (New Objective) and an applied spray voltage of 2.2 kV. The capillary temperature was set at 275°C. Full mass scan was acquired with mass range 375-1200 m/z in profile mode with resolution of 70000. The filling time was set at maximum of 250 ms with a limitation of $3x10^6$ ions. Data dependent acquisition (DDA) was performed with the resolution of the Orbitrap set to 35,000, with a fill time of 120 ms and a limitation of $2x10^5$ ions. A normalized collision energy of 32 was applied. Dynamic exclusion time of 30 s was used. The peptide match algorithm was set to ‘preferred’ and charge exclusion ‘unassigned’, charge states 1, 5 - 8 were excluded. MS^2 data was acquired in profile mode.

3.3.12 MS data analysis *

IsobarQuant (Franken *et al.*, 2015) and Mascot (v2.2.07) were chosen for data processing. A Uniprot *Saccharomyces cerrevisiae* proteome database (UP000002311) containing common contaminants and reversed sequences was used. The search parameters were the following: Carbamidomethyl (C) and TMT10 (K) (fixed modification), Acetyl (N-term), Oxidation (M) and TMT10 (N-term) (variable modifications). A mass error tolerance of 10 ppm was set for the full scan (MS1) and for MS/MS (MS2) spectra of 0.02 Da. Trypsin

* This method part was provided by Mandy Rettel and Frank Stein, MProteomics Core Facility, EMBL, Heidelberg, Germany.

was selected as protease with an allowance of maximum two missed cleavages. A minimum peptide length of seven amino acids and at least two unique peptides were required for a protein identification. The false discovery rate on peptide and protein level was set to 0.01.

3.3.13 Small scale fermentation of NWM for transcriptomics and proteomics

A single colony of the parental strain, two evolved isolates originating from Evolution niche E and two evolved isolates originating from Evolution niche G were grown overnight in 50 mL Falcon® tubes with 15 mL of YPD. The overnight cultures were inoculated in 50mL flasks that were filled up completely with natural wine must and were grown statically to recreate anaerobic conditions. The flasks were capped with autoclaved rubber plugs which were pierced with a small needle to allow the release of CO_2. The growth stage of the cultures was estimated based on weight loss which correlates to the consumption of glucose and release of CO_2. For this reason, the initial weight of the cultures was measured and followed once every day until no more weight loss was observed, at which point the cultures had entered the stationary phase. After the establishment of the growth kinetics with weight loss, the same cultures as described above were prepared, weight loss was once again followed and the cells were harvested at mid exponential phase for RNA-sequencing and proteomics analysis, as described in the previous sections.

3.3.14 Small scale fermentations in natural wine must (microvinification)

Natural wine must (NWM) was provided by Dr. Ramon Gonzalez (Instituto de Ciencias de la Vid y del Vino, Logrono, Spain) and cell growth was estimated based on weight loss/CO_2 release, as suggested by Harsch *et al.* (2009). In short, overnight cultures of yeast from YPD were washed 3 times with PBS and diluted to an initial OD_{600} of 0.1 in 55 mL of NWM. For the microvinification process, we used 50 mL Eelenmeyer flaks, which were filled to the maximum with the 55 mL of NWM, in order to create microanaerobic conditions. Maintaining the anaerobic conditions meant that the growth could not be estimated based on changes in the optical density, but it was correlated with the observed weight loss, which occurs from the release of CO_2, the end product of carbon metabolism. Release of CO_2 is possible through a small needle which is pierced through rubber plugs, which in turn are sterilised and used to seal the Eelenmeyer flaks, while a small piece of gauge prevents anything from the environment to fall inside the flask through the needle.

Contributions

- Dimitrios Konstantinidis (Structural and Computational Biology Unit, EMBL Heidelberg, Patil Group) designed and conducted the experiments, helped with the data analysis, visualisation and interpretation.
- Paula Jouhten (Structural and Computational Biology Unit, EMBL Heidelberg, Patil Group, VTT, Technical Research Center of Finland) designed the *EvolveX* algorithm and made the initial predictions for the evolution niches. Performed the whole genome data analysis. She was also responsible for overall data analysis and interpretation.
- Filipa Pereira (Structural and Computational Biology Unit, EMBL Heidelberg, Patil Group) performed analysis of the Transcriptomics data, and helped with experimental design.
- Kristina Grkovska (Structural and Computational Biology Unit, EMBL Heidelberg, Patil Group) helped collect samples for omics analysis.
- Rajna Hercog and Laura Villacorta (Genomics Core Facility, EMBL Heidelberg) performed the whole genome and the transcriptome sequencing.
- Mandy Rettel and Frank Stein (Proteomics Core Facility, EMBL Heidelberg) performed the proteomics sample preparation and initial data analysis.
- Ramon Gonzalez and Pillar Morales (ICVV, Instituto de Ciencias de la Vid y del Vino, Logrono, Spain) performed the small-scale wine fermentations and GCMS analysis of volatile aroma compounds.
- Kiran Raosaheb Patil (Structural and Computational Biology Unit, EMBL Heidelberg, Patil Group) supervised the project, designed experiments and helped with data interpretation.

Concluding Remarks

Microorganisms have been used by mankind throughout the centuries for a variety of applications. Use of microorganisms expanded rapidly through the advances of biotechnology and the new methodologies that emerged, resulted into the improvement of a great numbers of microbial species. One of these methodologies is ALE which has been used widely both to elucidate open questions about the microbial physiology and ecology, as well as for strain improvement. However, use of ALE remains limited to growth-associated traits like utilization of a difficult-to-assimilate nutrients, while its outcome cannot be fully directed towards a specific goal. In the present PhD Thesis, we described two novel ALE methodologies that can be used to overcome this limitation. The first one of the suggested methods, termed *Syn-A*, takes advantage of metabolic cross-feeding interactions between different species, to couple the production and secretion of a target compound with the survival and proliferation of the community. Increased concentration of the target compound in the environment would theoretically result into improved community fitness, despite of potentially high production costs for the secreting cells. We succeeded in proving the validity of this theory by constructing small mutualistic communities between LAB and the yeast *S. cerevisiae* for the increased production of B group vitamins, namely riboflavin and folate. Through the evolution process the bacterial cells increased the production and secretion of riboflavin up to 5-fold compared to the parental strain, by adapting their metabolism towards the increased production of the necessary for the riboflavin biosynthetic pathway precursors. Moreover, all the evolved strains that were tested exhibited upregulation of the enzymes that participate in the same pathway, suggesting that different aspects of the metabolism had to be regulated, in order to increase the production of the vitamin. The increase in production is specific and was observable only when we performed the evolution process in co-cultures and not in monoculture, while the outcome of the process was specific leading to increased production of the target vitamin.

The second method we described was the algorithm *EvolveX* which uses genome-scale metabolic model simulations to create the appropriate environment and hence selection pressure on specific metabolic fluxes, towards the production of a target compound. We proved the efficiency of the algorithm by evolving a wine yeast strain towards the increased production of organic compounds which affect the aroma profile of wine, but are not directly linked to survival and growth under normal environmental conditions. As in the case of *Syn-A* the evolution process with use of *EvolveX* is specifically directed towards the desired

outcome, since the evolved strains exhibited clear phenotypic improvements while retaining their ability to ferment wine must.

At first glance the importance of both of these novel methods lies in the possibility to improve specific phenotypic characteristics of industrially important strains for non-GMO applications, such as different food fermentations. For example, improved strains can be used to create affordable food products with increased nutritional value, which can help with the malnutrition that is observed in different ethnic and social groups of the human population. Alternatively, they can be used to produce products with improved sensory characteristics, like the improved aroma of fermented alcoholic drinks, which are highly desirable by consumers. Nevertheless, development of these two methodologies was not meant to be limited to food industry, but primarily they were developed as means to introduce precision in the field of laboratory evolution for all sorts of biotechnological and ecological applications.

List of Figures

List of Tables

Abbreviations

A.A	Amino acids
ABC	ATP binding cassette
ALE	Adaptive Laboratory Evolution
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
AUC	Area under the curve
Bp	Base pair
CDM	Chemically defined medium
COBRA	Constraint based reconstruction and analysis
DHF	Dihydrofolate
FAD	Flavin adenine dinucleotide
FBA	Flux balance analysis
FMN	Flavin mononucleotide
G6P	Glucose 6-phosphate
GEMs	Genome-scale metabolic models
GMO	Genetically Modified Organisms
GMP	Guanosine monophosphate
GO	Gene ontology
GRAS	Generally recognised as safe
GTP	Guanosine triphosphate
INDELs	Insertions and deletions
IMP	Inosine monophosphate
LAB	Lactic acid bacteria
LCMS	Liquid chromatography mass spectrometry
MarR	Multiple antibiotic resistance regulator
MATH	Microbial adhesion to hydrocarbons
MGAM	Modified Gifu Anaerobic Medium
MILP	Mixed-integer linear programming
MRS	De Man Rogosa and Sharpe medium
NAD	Nicotinamide adenine dinucleotide
NGS	Next generation sequencing
NWM	Natural wine must
OD	Optical density
pABA	p-amino benzoic acid
PAS domain	Per (period circadian protein) Arnt (aryl hydrocarbon receptor nuclear translocator protein) Sim (single-minded protein)
PBS	Phosphate buffered saline
PPP	Pentose phosphate pathway
PRPP	Phosphoribosyl pyrophosphate
PT	phosphotransferase
PTS	Phosphoenolpyruvate group translocation
Pur	purines
Pyr	pyrimidines
RDI	recommended daily intake
SNP	single nucleotide polymorphism
SWM	Synthetic wine must
TCA cycle	Tricarboxylic acid cycle

TetR	Tet repressor
THF	Tetrahydrofolate
UDP	Uridine diphosphate
UMP	Uridine monophospahte
UPLC	Ultra performance liquid chromatography
YPD	Yeast-extract Peptone Dextrose
WHO	World health organisation
WMM	Wine must mimicking medium

Literature

Abbas CA. and Sibirny AA., 2011. Genetic control of biosynthesis and transport of riboflavin and Flavin nucleotides and construction of robust biotechnological producers. Microbiol. Mol. Biol. Rev., 75(2):321-60.

Acevedo-Rocha CG., Gronenberg LS., Mack M., Commichau FM., Genee HJ., 2019. Microbial cell factories for the sustainable manufacturing of B vitamins. Curr. Opin. Biotechnol., 56:18-29.

Anders S., Pyl PT., Huber W., 2015. HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics, 31(2):166-9.

Bacher A., Eberhardt S., Fischer M., Kis K., Richter G., 2000. Biosynthesis of vitamin b2 (riboflavin). Annu Rev Nutr. 20:153-67.

Barve A. and Wagner A., 2013. A latent capacity for evolutionary innovation through exaptation in metabolic systems. Nature, 500(7461):203-6.

Belda I., Ruiz J., Esteban-Fernández A., Navascués E., Marquina D., Santos A., Moreno-Arribas V., 2017. Microbial Contribution to Wine Aroma and Its Intended Use for Wine Quality Improvement. Molecules, 22(189): 22020189.

Bennet A. and Hughes B., 2009. Microbial experimental evolution. Am. J. Physiol. Regul. Integr. Comp. Physiol., 297: R17-R25.

Boon E., Meehan C., Whidden C., Wong D., Langille M., Beiko R., 2014. Interactions in the microbiome: communities of organisms and communities of genes. FEMS Microbiol. Rev., 38: 90-118.

Borneman A., Desany B., Riches D., Affourtit J., Forgan A., Pretorius I., Egholm M., Chambers P., 2011. Whole-Genome Comparison Reveals Novel Genetic Elements That Characterize the Genome of Industrial Strains of *Saccharomyces cerevisiae*. PLoS Genet. 7(2): e1001287.

Boulton SJ. And Jackson Sp., 1996. *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as barrier to error-prone DNA repair pathways. EMBO J., 15:5093-5103.

Bravo D., Landete J., 2017. Genetic engineering as a powerful tool to improve probiotic strains. Biotechnology and Genetic Engineering Review, 33(2).

Brosius J., 2019. Exaptation at the molecular genetic level. Sci. China Life Sci., 62(4):437-452.

Burgess C., O'Connell-Motherway M., Sybesma W., Hugenholtz J., van Sinderen D., 2004. Riboflavin production in Lactococcus lactis: potential for in situ production of vitamin-enriched foods. Appl. Environ. Microbiol., 70(10):5769-77.

Burgess C., Smid E., Rutten G., van Sinderen D., 2006. A general method for selection of riboflavin-overproducing food grade micro-organisms. Microbial Cell Factories, 5:24.

Burgess C., Smid E. and van Sinderen D., 2009. Bacterial vitamin B2, B11 and B12 overproduction: An overview. Int. J. Food Microbiol., 133(1-2):1-7.

Butel MJ., 2014. Probiotics, gut microbiota and health. Médecine et Maladies Infectieuses, 1:1-8.

Capozzi V., Russo P., Dueñas MT., López P, Spano G., 2012. Lactic acid bacteria producing B-group vitamins: a great potential for functional cereal products. Appl. Microbiol. Biotechnol., 96(6):1383-94.

Carrizo S., LeBlanc A., LeBlanc J., Rollán G., 2020. Quinoa pasta fermented with lactic acid bacteria prevents nutritional deficiencies in mice. Food Research International, 127: 108735.

Chen J., Vestergaard M., Glasdam T., Shen J., Dufva M., Solem C., Jensen P., 2017. Finding the Needle in the Haystack – the Use of Microfluidic Droplet Technology to Identify Vitamin-Secreting Lactic Acid Bacteria. mBio, 8: e00526-17.

Chen H., Wu C. and He X., 2018. The Genotype-Phenotype Relationships in the Light of Natural Selection. Molecular Biology and Evolution, 35(3):525-542.

Claesson M., van Sinderen D., O'Toole P., 2007. The genus *Lactobacillus*- a genomic basis for understanding its diversity. FEMS Microbiol. Lett., 269:22-28.

Conrad T., Lewis N. and Palsson B., 2011. Microbial laboratory evolution in the era of genome-scale science. Molecular Systems Biology, 7:509.

Curiel J., Salvadó Z., Tronchoni J., Morales P., Rodrigues A., Quirós M., Gonzalez R., 2016. Identification of target genes to control acetate yield during aerobic fermentation with *Saccharomyces cerevisiae*. Microb. Cell Fact., 15:156.

Cuthbertson L. and Nodwell JR., 2013. The TetR family of regulators. Microbiol. Mol. Biol. Rev., 77(3): 440-75.

Darzi Y., Letunic I., Bork P., Yamade T., 2018. iPath3.0: interactive pathways explorer v3. Nucleic Acids Research, 46(1):510-13.

Dayon L., Hainard A., Licker V., Turck N., Kuhn K., Hochstrasser D., Burkhard P., Sanchez J., 2008. Relative Quantification of Proteins in Human Cerebrospinal Fluids by MS/MS using 6-Plex Isobaric Tags. Anal. Chem., 80(8):2921-2931.

Deathrage D., Barrick J, 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol. Biol. 1151:165-188.

Deathrage D., Kepner J., Bennett A., Lenski R., Barrick J., 2017. Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures. PNAS, 114(10): 1904-1912.

Deepika G., Charalampopoulos D., 2010. Surface and adhesion properties of lactobacilli. Adv. Appl. Microbiol., 70: 127-52.

de Vos WM. and Hugenholtz J., 2004. Engineering metabolic highways in Lactococci and other lactic acid bacteria. Trends Biotechnol., 22(2):72-9.

Dietrich J., McKee A. and Keasling J., 2010. High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection. Annual Review of Biochemistry, 79:563-590.

Dobin A., Davis C., Schlesinger F., Drenkow J., Zaleski C., Jha S., Batut P., Chaisson M., Gingeras T., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics, 29(1):15-21.

Dobzhansky T., 1973. Nothing in Biology Makes Sense except in the light of Evolution. The American Biology Teacher, 35(3):125-129.

Dragosits M. and Mattanovich D., 2013. Adaptive laboratory evolution-principles and applications for biotechnology. Microbial Cell Factories, 12:64.

D'Souza G., Waschina S., Pande S., Bohl K., Kaleta C., Kost C., 2014. Less is more: selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. Evolution, 68(9): 2559-70.

D'Souza G., Shitut S., Preussger D., Yousif G., Waschina S., Kost C., 2018. Ecology and evolution of metabolic cross-feeding interactions in bacteria. Nat Prod Rep., 35(5): 455-488.

Fernández M., Zúñiga M., 2006. Amino acid catabolic pathways of lactic acid bacteria. Crit. Rev. Microbiol., 32(3):155-83.

Food and Agricultural Organization of the United Nations and World Health Organization. Health and nutritional properties of probiotics in food including powder milk with live lactic acid bacteria. *World Health Organization.*

Franken H., Mathleson T., Childs D., Sweetman GM., Werner T., Tögel I., Doce C., Gade S., Bantscheff M., Drewes G., Reinhard FB., Huber W., Savitski M., 2015. Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. Nat. Protoc. 10(10):1567-93.

Fugelsang K.C., Edwards CG., 2007. Wine Microbiology: Practical Applications and Procedures. Springer, New York.

Garcia V., Rivera J., Contreras A., Ganga M., Martinez C., 2012. Development and characterization of hybrids from native wine yeasts. Braz. J. Microbiol., 43(2):482-489.

Ghosh T., Beniwal A., Semwal A., Navani N., 2019. Mechanistic Insights Into Probiotic Properties of Lactic Acid Bacteria Associated With Ethnic Fermented Dairy Products. Front. Microbiol., 10:502.

Gibson B., Wilson D., Feli E. and Eyre-Walker A., 2018. The distribution of bacterial doubling times in the wild. Proc. R. Soc. B, 285:20180789.

Giraud A., Matic I., Tenaillon O., Clara A., Radman M., Fons M., Taddei F., 2001. Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. Science, 291(5513):2606-8.

Goh YJ., Barrangou R., 2019. Harnessing CRISPR-Cas systems for precision engineering of designer probiotic lactobacilli. Curr Opin Biotechnol., 56:163-171.

Goldstein A.L. and McCusker J.H., 1999. Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. Yeast, 15(14):1541-53.

Gonzalez R. and Morales P., 2017. Wine secondary aroma: understanding yeast production of higher alcohols. Microb. Biotechnol., 10(6):1449-1450.

Grand View Research, 2019. Functional Food Market Analysis, 2019. Grand View Research, Inc., USA (https://www.grandviewresearch.com).

Gresham D. and Hong J., 2015. The functional basis of adaptive evolution in chemostats. FEMS Microbiol. Rev., 39(1):2-16.

Güldener U., Heck S., Fielder T., Beinhauer J., Hegemann J.H., 1996. A new efficient gene disruption cassette for repeated use in budding yeast. Nucleic Acids Res., 24(13):2519-24.

Guo J., Brosnan B., Furey A., Arendt E., Murphy P., Coffey A., 2012. Antifungal activity of Lactobacillus against *Microsporum canis*, *Microsporum gypseum* and *Epidermophyton floccosum*. Bioeng. Bugs, 3(2):102-111.

Gupta R., 2016. Impact of genomics on the understanding of microbial evolution and classification: the importance of Darwin's views on classification. FEMS Microbiology Reviews, 40:520-553.

Harcombe W., 2010. Novel Cooperation Experimentally Evolved Between Species. Evolution, 64(7):2166-2172.

Harcombe W., Delaney N., Leiby N., Klitgord N., Marx C., 2013. The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum. PLoS Comput. Biol., 9(6): e1003091.

Harcombe W., Chacón J., Adamowicz E., Chubiz L., Marx C., 2018. Evolution of bidirectional costly mutualism from byproduct consumption. PNAS, 115(47):12000-12004.

Hazelwood L.A., Daran J.M., van Maris J.A., Pronk J.T., Dickinson J.R., 2008. The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism. Appl. Environ. Microbiol., 74, 2259–2266.

Heirendt L., Arreckx S., Pfau T., Mendoza S.N., Richelle A., Heinken A., Haraldsdóttir H.S., Wachowiak J., Keating S.M., Vlasov V., Magnusdóttir S., Ng SY., Preciat G., Žagare A., Chan SHJ., Aurich MK., Clancy CM., Modamio J., Sauls JT., Noronha A., Bordbar A., Cousins B., El Assal DC., Valcarcel LV., Apaolaza I., Ghaderi S., Ahookhosh M., Ben Guebila M., et al., 2019. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat. Protoc., 14(3):639-702.

Hillesland K., 2018. Evolution on the bright side of life: microorganisms and the evolution of mutualism. Ann. N. Y. Acad. Sci., 1422(1): 88-103.

Hoek T., Axelrod K., Biancalani T., Yurtsev E., Liu J., Gore J., 2016. Resource availability modulates the cooperative and competitive nature of a microbial cross-feeding mutualism. PLOS Biology, 14(8): e1002540.

Huang da W., Sherman B.T., Lempicki R.A., 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc., 4(1):44-57.

Hugenholtz J., Smid EJ., 2002. Nutraceutical production with food-grade microorganisms. Curr Opin Biotechnol., 13(5):497-507.

Hughes C., Moggridge S., Müller T., Sorensen P., Morin G., Krijgsveld J., 2019. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. Nat. Protoc., 14:68-85.

Iyer R. and Tomar SK., 2009. Folate: A functional Food Constituent. Journal of Food Science, 74(9): 114-122.

Jagmann N. and Philipp B., 2014. Reprint of Design of synthetic microbial communities for biotechnological production processes. J. Biotechnol., 192: 293-301.

Ju SY., Kim JH., Lee PC., 2016. Long-term adaptive evolution of *Leuconostoc mesenteroides* for enhancement of lactic acid tolerance and production. Biotechnol Biofules, 9:240.

Koskella B., Vos M., 2015. Adaptation in Natural Microbial Populations. Annu. Rev. Ecol. Evol. Syst. 46:503-22.

Kotzamanidis C., Kourelis A., Litopoulou-Tazanetaki E., Tzanetakis N., Yiangou M., 2010. Evaluation of adhesion capacity, cell surface traits and immunomodulatory activity of presumptive probiotic Lactobacillus strains. Int. J. Food Microbiol., 140(2-3_:154-63.

Kussel E., 2013. Evolution in Microbes. Ann. Rev. Biophys. 42:493-514.

LaCroix R., Palsson B., Feist A., 2017. A Model for Desogning Adaptive Laboratory Evolution Experiments. Appl. Environ. Microbiol. 83(8): e03115-16.

Laiño JE., del Valle M., de Giori G., LeBlanc J., 2013. Development of a high folate concentration yoghurt naturally bio-enriched using selected lactic acid bacteria. LWT-Food Scienc. and Technol., 54(1):1-5.

Lang G., Desai M., 2014. The spectrum of adaptive mutations in experimental evolution. Genomics, 104(0):412-416.

LeBlanc JG., Burgess C., Sesma F., de Giori S., van Sinderen D., 2005. Ingestion of Milk Fermented by Genetically Modified *Lactococcus lactis* Improves the riboflavin Status of Deficient Rats. J. Dairy Sci., 88:3435-3442.

LeBlanc JG., Laiño JE., del Valle MJ., Vannini V., van Sinderen D., Taranto MP., de Valdez GF., de Giori GS., Sesma F., 2011. B-group vitamin production by lactic acid bacteria-current knowledge and potential applications. J. Appl. Microbiol., 111(6):1297-309.

LeBlanc JG., Milani C., de Giori GS., Sesma F., van Sinderen D., Ventura M., 2013. Bacteria as vitamin suppliers to their host: a gut microbiota perspective. Curr. Opin. Biotechnol., 24(2): 160-8.

LeBlanc JG., Levit R., de Giori G., LeBlanc A., 2020. Application of vitamin-producing lactic acid bacteria to treat intestinal inflammatory diseases. Appl. Microbiol. Biotechnol., 104(8):3331-3337.

Lee S., Mattanovich D. and Villaverde A., 2012. Systems metabolic engineering, industrial biotechnology and microbial cell factories. Microbial Cell Factories, 11:156.

Lenski R., Rose M., Simpson S and Tadler S., 1991. Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2.000 Generations. The American Naturalist, 138(6):1315-1341.

Lenski R., 2017. Experimental evolution and the dynamics of adaptation and genome evolution in microbial populations. ISME J., 11(10):2181-2194.

Levit R., Savoy de Giori G., LeBlanc A., LeBlanc JG., 2017. Effect of riboflavin-producing bacteria against chemically induced colitis in mice. Journal of Applied Microbiology, 124(1):232-240.

Lewis N., Nagarajan H., Palsson B., 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of *in silico* methods. Nat. Rev. Microbiol., 10(4):291-305.

Li H. and Durbin R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14):1754-60.

Libby E., Hébert-Dufresne, Hosseini S., Wagner A., 2019. Syntrophy emerges spontaneously in complex metabolic systems. PLoS Comput. Biol., 15(7): e1007169.

Lindermann S., Bernstein H., Song HS., Fredrickson J., Fields M., Shou W., Johnson D., Beliaev A., 2016. Engineering microbial consortia for controllable outputs. The ISME Journal, 10: 2077-2084.

Lo C. and Chain P., 2014. Rapid evaluation and quality control of next generation sequencing data with FaQCs. BMC Bioinformatics, 15(1):366.

Lopes H. and Rocha I., 2017. Genome-scale modeling of yeast: chronology, applications and critical perspectives. FEMS Yeat Res., 17(5): fox050.

Lourens-Hattingh, A. and Viljoen, B.C., 2001. Yoghurt as probiotic carrier food. International Dairy Journal, 11:1–17.

Love M., Huber W., Anders S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology, 15:550.

Lu H, Feiran L., Benjamin JS., Zhengming Z., Gang L., Domenzain I., Marcišauskas S., Anton PM., Lappa D., Lieven C., Beber ME., Sonnenschein N., Kerkhoven E., Nielsen J., 2019. A consesnsus *S. cerevisiae* metabolic model Yeast8 and its ecosystems for comprehensively probing cellular metabolism. Nat. Commun., 10:3586.

Machado D., Andrejev S., Tramontano M., Patil KR., 2018. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. Nucleic Acids Research, 46(15):7542-7553.

Marchal M., Goldschmidt F., Derksen-Müller S., Panke S., Ackermann M., Johnson R., 2017. A passive mutualistic interaction promotes the evolution of spatial structure within microbial populations. BMC Evol. Biol., 17: 106.

Marsit S., Dequin S., 2015. Diversity and adaptive evolution of *Saccharomyces* wine yeast: a review. FEMS Yeast Res., 15(7): fov067.

Martin M., 2011. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. EMBnet.journal, 17(1).

McCully A., LaSarre B., McKinlay J., 2017. Recipient-Biased Competition for an Intracellularly Generated Cross-Fed Nutrient Is Required for Coexistence of Microbial Mutualists. mBio, 8(6): e01620-17.

McDonald M., 2019. Microbial Experimental Evolution – a proving ground for evolutionary theory and a tool for discovery. EMBO Reports, 20: e46992.

Mee M., Collins J., Church G., Wang H., 2014. Syntrophic exchange in synthetic microbial communities. PNAS, 111(20):E2149-E2156.

Merritt JH., Kadouri DE., O'Toole GA, 2005. Growing and analysing static biofilms. Curr. Protoc. Microbiol., Chapter1: Unit 1B.1.

Meucci A., Rossetti L., Zago M., Monti L., Giraffa G., Carminati D., Tidona F., 2018. Folates biosynthesis by *Streptococcus thermophiles* during growth in milk. Food Microbiol., 69: 116-122.

Mülleder M., Capuano F., Pir P., Christen S., Sauer U., Oliver S., Ralser M., 2012. A prototrophic deletion mutant collection for yeast metabolomics and systems biology. Nat. Biotechnol., 30(12):1176-1178.

National Academy of Sciences; Avise JC, Ayala FJ (Editors), 2007. In the Light of Evolution: Volume I: Adaptation and Complex Design. Washington (DC): National Academies Press (US).

Nissan T., Galani K., Maco B., Tollervey D., Aebi U., Hurt E., 2004. A Pre-Ribosome with a Tadpole-like Structure Functions in ATP-Dependent Maturation of 60S Subunits. Molecular Cell, 15(2):295-301.

Nosil P., Villoutreix R., Carvalho C., Farkas T., Soria-Carrasco, Feder J., Crespi B., Gompert Z., 2018. Natural selection and the predictability of evolution in *Timena* stick insects. Science, 359(6377):765-770.

Novo M., Bigey F., Beyne E., Galeote V., Gavory F., Mallet S., Cambon B., Legras J.L. Wincjer P., Casaregole S., Dequin S., 2009. Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast Saccharomyces cerevisiae EC1118. Proc. Natl. Acad. Sci. USA, 106(38):16333-8.

Nurk S., Bankevich A., Antipov D., Gurevich A., Korobeynikov A., Lapidus A., Prjibelsky A., Pyshkin A., Sirotkin A., Sirotkin Y., Stepanauskas R., McLean J., Lasken R., Clingenpeel S., Tesler G., Alekseyev M., Pevzer P., 2013. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. Research in Computational Molecular Biology. RECOMB 2013. Lecture Notes in Computer Science, vol. 7821. Springer, Berlin, Heidelberg.

O'Brien EJ., Monk JM., Palsson BO., 2015. Using Genome-scale Models to Predict Biological Capabilities. Cell, 161(5):971-987.

Österlund T., Nookaew I., Nielsen J., 2012. Fifteen years of large scale metabolic modeling of yeast: developments and impacts. Biotechnol. Adv., 30(5):979-88.

Otero J. and Nielsen J., 2010. Industrial Systems Biology. Biotechnol. Bioeng., 105: 439-460.

Overbeek R., Olson R., Pusch G., Olsen G., Davis J., Disz T., Edwards R., Gerdes S., Parrello B., Shukla M., Vonstein V., Wattam A., Xia F., Stevens R., 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic. Acids Res., 42:D206-D214.

Pandal J. and Noirel J., 2014. Synthetic microbial ecosystems for biotechnology. Biotchnol. Lett., 36(6): 1141-51.

Panigrahi P., Parida S., Nanda N., Satpathy R., Pradhan L., Chandel D., Baccaglini L., Mohapatra A., 2017. A randomized symbiotic trial to prevent sepsis among infants in rural India. Nature, 548:407-412.

Papagianni M., 2012. Metabolic engineering of lactic acid bacteria for the production of industrially important compounds. Comput. Struct. Biotechnol. J., 3: e201210003.

Pande S., Merker H., Bohl K., Reichelt M., Schuster S., Figueiredo L., Kaleta C., Kost C., 2014. Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. ISME Journal, 8:953-962.

Pande S. and Kost C., 2017. Bacterial Unculturability and the Formation of Intercellular Metabolic Networks. Trends in Microbiology, 25(5): 349-361.

Plaut G.E., 1954. Biosynthesis of Riboflavin II. Incorporation of ^{14}C-labelled compounds into ring. A. J. Biol. Chem., 208:513-520.

Ponomarova O., Gabrielli N., Sévin DC., Mülleder M., Zirngibl K., Bulyha K., Andrejev S., Kafkia E., Typas A., Sauer U., Rasler M., Patil KR., 2017. Yeast Creates a Niche for Symbiotic Lactic Acid Bacteria through Nitrogen Overflow. Cell Syst., 5(4):345-357.

Portnoy V., Bezdan D., Zengler K., 2011. Adaptive laboratory evolution – harnessing the power of biology for metabolic engineering. Current Opinion in Biotechnology, 22:590-594.

Poursat B., van Spanning R., Voogt P., Parsons J., 2019. Implications of microbial adaptation for the assessment of environmental persistence of chemicals. Critical Reviews in Environmental Science and Technology, 49(23):2220-2255.

Rainey PB., Remigi P., Farr A., Lind P., 2017. Darwin was right: where now for experimental evolution? Current Opinion in Genetics & Development. 47:102-109.

Regulation (EC) No 1829/2003 of the European Parliament and of the Council of 22 September 2003 on genetically modified food and feed. Official Journal L 268, 18/10/2003 P. 0001-0023.

Reichel M., Liao Y., Rettel M., Ragan C., Evers M., Alleaume A., Horos R., Hentze M., Preiss T., Millar A., 2016. In Planta Determination of the mRNA-Binding Proteome of Arabidopsis Etiolated Seedings. Plant Cell, 28:2435-2452.

Richardson E., Zerr K., Tsaousis A., Dorrell R. and Dacks J., 2015. Evolutionary cell biology: functional insight from "endless forms most beautiful". Mol. Biol. Cell, 26(25):4532-4538.

Roff D., Heibo E., Vøllestad, 2006. The importance of growth and mortality costs in the evolution of the optimal life history. Journal of Evolutionary Biology, 19(6): 1920-1930.

Rossi M., Amaretti A., Raimondi S., 2011. Folate production by probiotic bacteria. Nutrients, 3(1):118-34.

Russo, P., Capozzi, V., Arena, M.P., Spadaccino, G., Duenas, M.T., Lopez, P., *et al*. (2014) Riboflavin-overproducing strains of *Lactobacillus fermentum* for riboflavin-enriched bread. *Appl Microbiol Biotechnol* 98**:** 3691–3700.

Salari R. and Salari R., 2017. Investigation of the Best *Saccharomyces cerevisiae* Growth Condition. Electronic Physician, 9(1):3592-3597.

Sambamoorthy G., Sinha H., Raman K., 2019. Evolutionary design principles in metabolism. Proc. R. Soc. B., 286: 20190098.

Sandberg T., Lloyd CJ., Palsson B., Feist AM., 2017. Laboratory Evolution to Alternating Substrate Environments Yields Distinct Phenotypic and Genetic Adaptive Strategies. Appl. Environ. Microbiol., 83(13): e00410-17.

Saubade F., Hemery Y., Guyot JP., Humblot C., 2017. Lactic acid fermentation as a tool for increasing the folate content of foods. Cr. Rev. in Food Scien. and Nutr, 18: 3894-3910.

Sauer M., Russmayer H., Grabherr R., Peterbauer CK., Marx H., 2017. The Efficient Clade: Lactic Acid Bacteria for Industrial Chemical Production.

Schindelin J., Arganda-Carreras I., Frise E., Kaynig V., Longair M., Pietzsch T., Preibisch S., Rueden C., Saalfeld S., Schmid B., Tinevez JY., White DJ., Hartenstein V., Eliceiri K., Tomancak P., Cardona A., 2012. Fiji: an open-source platform for biological-image analysis. Nature Methods, 9:676-682.

Schlüpen C., Santos M.A., Weber U., de Graaf A., Revuelta J.L., Stahmann K.P., 2003. Disruption of the SHM2 gene, encoding one of two serine hydroxymethyltransferase isoenzymes, reduces the flux from glycine to serine in *Ashbya gossypii*. Biochem J., 369:263–273.

Schmidt-Dannert C., Arnold FH., 1999. Directed evolution of industrial enzymes. Trends Biotechnol., 17(4):135-6.

Schwechheimer SK., Park EY., Revuelta JL., Becker J., Wittmann C., 2016. Biotechnology of riboflavin. Appl. Microbiol. Biotechnol., 100(5):2107-19.

Shi T., Wang Y., Wang Z., Wang G., Liu D., Fu J., Chen T., Zhao X., 2014. Deregulation of purine pathway in *Bacillus subtilis* and its use in riboflavin biosynthesis. Microbial Cell Factories, 13:101.

Shlomi T., Berkman O., Ruppin E. 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. P Natl Acad Sci USA, 102:7695–7700.

Silva R., Aguiar T., Domingues L., 2015. Blockage of the pyrimidine biosynthetic pathway affects riboflavin production in *Ashbya gossypii*. J. Biotechnol., 193: 37-40.

Silva R., Aguiar T., Oliveira C., Domingues L., 2019. Physiological characterization of a pyrimidine auxotroph exposes link between uracil phosphoribosyltransferase regulation and riboflavin production in *Ashbya gossypii*. New Biotechnology, 50:1-8.

Sjostrom S., Huang M., Nielsen J., Joensson H., Svahn H., 2014. Micro-droplet based directed evolution outperforms conventional laboratory evolution. International conference on Miniaterized Systems for Chemistry and Life Sciences (proceedings), 169-171.

Smid EJ. And Kleerebezem M., 2014. Production of aroma compounds in lactic fermentations. Annu. Rev. Food Sci. Technol., 5:313-26.

Smith N., Shorten P., Altermann E., Roy N., McNabb W., 2019. The Classification and Evolution of Bacterial Cross-Feeding. Front. Ecol. Evol., 7: 153.

Steensels J., Snoek T., Meersman E., Picca Nicolino M., Voordeckers K., Versterpen KJ., 2014. Improving industrial yeast strains: exploiting natural and artificial diversity. FEMS Microbiol. Rev., 38(5):947-95.

Steinmetz L.M., Scharfe C., Deutschbauer A.M., Mokranjac D., Herman Z.S., Jones T., Chu A.M., Giaever G., Prokisch H., Oefner P.J., Davis R.W., 2002. Systematic screen for human disease genes in yeast. Nat. Genet., 31(4):400-4.

Stellwagen A., Haimberger Z., Veatch J., Gottschling D., 2003. Ku interacts with telomerase RNA to promote telomere addition at native and broken chromosome ends. Genes and Development, 17:2384-2395.

Stephanopoulos G., 2012. Synthetic biology and metabolic engineering. ACS Synth. Biol., 1(11):514-25.

Stovicek V., Borja G.M., Forster J., Borodina I., 2015. EasyClone 2.0: expanded toolkit of integrative vectors for stable gene expression in industrial *Saccharomyces cerevisiae* strains. J. Ind. Microbiol. Biotechnol., 42(11):1519-31.

Strimmer K., 2008. A unified approach to false discovery rate estimation. BMC Bioinformatics, 9:303.

Strucko T., Zirngibl K., Pereira F., Kafkia E., Mohamed ET., Rettel M., Stein F., Feist AM., Jouhten P., Patil KR., ForsterJ., 2018. Laboratory evolution reveals regulatory and metabolic trade-offs of glycerol utilization in *Saccharomyces cerevisiae*. Metab. Eng., 47:73-82.

Stump S., Johnson E.C., Klausmeier C., 2018. Local interactions and self-organized spatial patterns stabilize microbial cross-feeding against cheaters. J. R. Soc. Interface 15: 20170822.

Tenaillon O., Barrick J., Ribeck N., Deathrage D., Blanchard J., Dasgupta A., Wu G., Wilgoss S., Cruveiller S., Médigue C., Schneider D., Lenski R., 2016. Tempo and mode of genome evolution in a 50.000-generation experiment. Nature, 536(7615):165-70.

Tenaillon O., 2018. Experimental evolution heals the scars of genome-scale recoding. PNAS, 115(12):2853-2855.

Teusenik B. and Smid EJ., 2016. Modeling strategies for the industrial exploitation of lactic acid bacteria. Nat. Rev. Mivrobiol., 4(1):46-56.

Thakur K., Tomar SK. And De S., 2016. Lactic acid bacteria as a cell factory for riboflavin production. Microb. Biotechnol., 9(4): 441-451.

Trausch J., Ceres P., Reyes F., Batey R., 2011. The structure of a tetrahydrofolate sensing riboswitch reveals two ligand binding sites in a single aptamer. Structure, 19(10): 1413-1423.

Turkarslan S., Raman A., Thompson A., Arens C., Gillespie M., von Netzer F., Hillesland K., Stoylar S., de Lomana A., Reiss D., Gorman-Lewis G., Ranish J., Wall J., Stahl D., Baliga N., 2017. Mechanism for microbial population collapse in a fluctuating resource environment. Molecular Systems Biology, 13(3):919.

van den Bergh B., Swings T., Fauvart M., Michiels J., 2018. Experimental Design, Population Dynamics, and Diversity in Microbial Experimental Evolution. Microbiology and Molecular Biology Reviews, 82(3):1-55.

Von Canstein H., Ogawa J., Shimizu S., Lloyd JR., 2008. Secretion of flavins by Shewanella species and their role in extracellular electron transfer. Appl. Environ. Microbi., 74(3):615-23.

Wallace AR., 1889. Darwinism, 1st edn (London: Macmillan).

Wegkamp A., van Oorschot W., de Vos WM. and Smid EJ., 2007. Characterization of the role of para-aminobenzoic acid biosynthesis in folate production by *Lactococcus lactis*. Appl. Enviro.n Microb., 73:2673–2681.

Werner T., Sweetman G., Savitski M., Mathieson T., Bantscheff M., Savitski M., 2014. Ion Coalescence of Neutron Encoded TMT 10-Plex Reporter Ions. Anal. Chem., 86(7):3594-3601.

Widyastuti Y., Rohmatussolihat, Febrisiantosa A., 2014. The Role of Lactic Acid Bacteria in Milk Fermentation. Food and Nutrition Sciences, 5:435-442.

Wildenberg G. and Murray A., 2014. Evolving a 24-hr oscillator in budding yeast. eLife, 3: e04875.

Winkler J., Reyes L., Kao K., 2013. Adaptive laboratory evolution for strain engineering. Methods in Molecular Biology, 985.

Woese C., 1987. Bacterial Evolution. Microbiological Reviews, 51(2):221-271.

Zhang J., Kobert K., Flouri T., Stamatakis A., 2014. PEAR: a fast accurate Illumina Paired-End reAd mergeR. Bioinformatics, 30(5):614-20.

Zhang X. and Reed J., 2014. Adaptive Evolution of Synthetic Cooperating Communities Improves Growth Performance. PLoS ONE, 9(10): e108297.

Ziesack M., Gibson T., Oliver J., Shumaker A., Hsu B., Riglar D., Giessen T., DiBenedetto N., Bry L., Way J., Silver P., Gerber G., 2019. Engineering Interspecies Amino Acid Cross-Feeding Increases Population Evenness in a Synthetic Bacterial Consortium. mSystems, 4: e00352-19.