Generation of bacterial platform strains for raw materials utilization using adaptive laboratory evolution

Mohamed, Elsayed Tharwat Tolba

Publication date:
2020

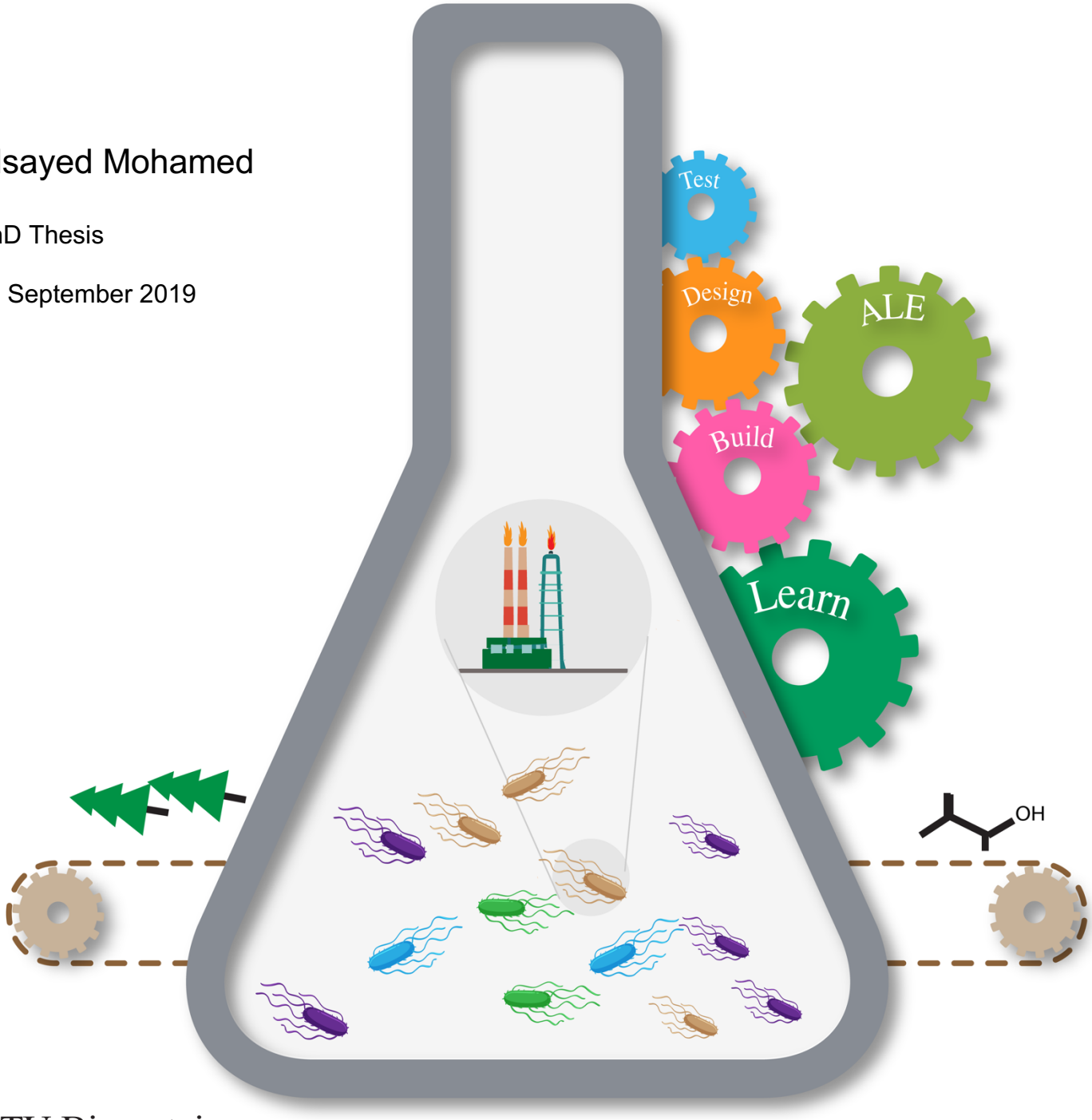
Document Version
Publisher's PDF, also known as Version of record

Generation of bacterial platform strains for raw materials utilization using adaptive laboratory evolution

Elsayed Mohamed

PhD Thesis

13 September 2019



DTU Biosustain
The Novo Nordisk Foundation Center for Biosustainability

Generation of bacterial platform strains for raw materials utilization using adaptive laboratory evolution

PhD Thesis by Elsayed Mohamed

Full name: Elsayed Tharwat Tolba Mohamed

13 September 2019

DTU Biosustain

The Novo Nordisk Foundation Center for Biosustainability

Generation of bacterial platform strains for raw materials utilization using adaptive laboratory evolution

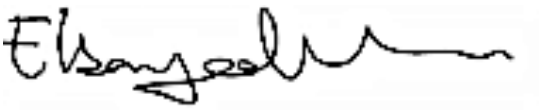
PhD Thesis by Elsayed Mohamed

Principal supervisor: Adam Feist

Co-supervisor: Markus Herrgård

Preface

This thesis is written as partial fulfillment of the requirements for obtaining a PhD degree at the Technical University of Denmark. The work included in the thesis has been carried out at the Novo Nordisk Foundation Center for Biosustainability in the period from June 2015 to September 2019 as part time Ph.D. program. The work has been supervised by group leader/Senior Researcher, Dr. Adam Feist and Professor Markus Herrgård. The entire work was funded by the Technical University of Denmark and the Novo Nordisk Foundation.



Elsayed Mohamed Kgs.

Lyngby,13 September 2019

Abstract

Rising global temperatures and limited fossil resources make it increasingly urgent to find alternative ways of producing fuels and chemicals. Metabolic engineering offers promising solutions to this problem by utilizing biological systems—microbes—as cell factories for manufacturing a diverse set of products from renewable resources. However, rationally developing a microbial cell factory requires an extensive amount of knowledge of cellular processes as well as expensive and time-consuming molecular biology to design strains with desired characteristics. Adaptive laboratory evolution (ALE) is an alternative approach that can accelerate the development of microbial cell factories by harnessing the power of evolutionary processes. An important aspect of sustainable bio-production is the utilization of renewable raw substrates, as they are abundant and have the potential to be a low-cost feedstock. However, utilization of raw materials can be limited by two issues: a) when deconstructed, they often contain toxic materials that affect cell fitness, and b) most microbial strains are limited in their consumption of such materials. It has been shown that ALE can be used to overcome both issues and, further, that understanding the genetic basis of the desired phenotypes is possible, given the availability of whole-genome sequencing. This thesis addresses the generation of platform strains optimized for the utilization and tolerance of raw material feedstocks, including lignocellulosic hydrolysates and raw sugar sources, focusing on an ALE approach. The main focus for raw materials was on lignocellulosic hydrolysate and raw sugarcane juice. For lignocellulosic hydrolysate, two studies were performed; the first was to overcome the toxicity of ionic liquid (a promising solvent for lignocellulosic biomass), and the second was to increase bacterial tolerance and utilization of the aromatic acids typically found in lignocellulosic hydrolysates, especially from solubilized lignin. A third study was also conducted to establish efficient utilization of sucrose (the main carbon of sugarcane juice) in industrial-relevant strains with improved overall fitness. This thesis contributes to our understanding of how microbial cells adapt to specific stress and growth conditions and provides tangible platform strains and a set of mutations that can be used as engineering tools to generate production strains for biomaterials based on renewable feedstocks.

Dansk resumé

Stigende globale temperaturer og begrænsede fossile ressourcer gør det mere og mere presserende at finde alternative måder at fremstille brændstoffer og kemikalier på. Metabolic engineering byder på lovende løsninger på dette problem, ved at anvende biologiske systemer - mikrober - som cellefabrikker til fremstilling af et forskelligartet sæt af produkter fra vedvarende ressourcer. Imidlertid kræver det at udvikle en mikrobiel cellefabrik, en omfattende mængde viden om cellulære processer, og det er dyrt og tidskrævende at designe stammer med de ønskede egenskaber vhja molekylær biologiske metoder. Adaptiv laboratorieudvikling (ALE) er en alternativ tilgang, der kan fremskynde udviklingen af mikrobielle cellefabrikker ved at udnytte de evolutionære processer. Et vigtigt aspekt ved bæredygtig bioproduktion er brugen af vedvarende råmaterialer, da de er rigelige og har potentialet til at være et billigt materiale. Anvendelse af råmaterialer kan imidlertid være begrænset af to problemer: a) når de dekonstrueres, indeholder de ofte toksiske materialer, der påvirker cellen, og b) de fleste mikrobielle stammer er begrænset i deres forbrug af sådanne materialer. Det er vist, at ALE kan bruges til at overvinde begge problemer, og endvidere at forstå det genetiske grundlag for de ønskede fænotyper er muligt i betragtning af tilgængeligheden af hel-genom-sekventering. Denne afhandling omhandler genereringen af platformstammer, der er optimeret til at udnytte og tolerere råmaterialer, herunder lignocellulosehydrolysater og råsukkerkilder, med fokus på en ALE-tilgang. Hovedfokus mht råmaterialer var lignocellulosic hydrolysat og sukkerrørssaft. For lignocellulosisk hydrolysat blev der udført to undersøgelser; den første var at overvinde toksiciteten af ionisk væske (et lovende opløsningsmiddel til lignocellulosisk biomasse), og det andet var at øge bakterietolerance og anvendelse af de aromatiske syrer, der typisk findes i lignocellulosiske hydrolysater, især fra solubiliseret lignin. En tredje undersøgelse blev også udført for at etablere effektiv anvendelse af saccharose (det vigtigste kulstof af sukkerrørsaft) i industrielle relevante stammer med forbedret generel egnethed. Denne afhandling bidrager til vores forståelse af, hvordan mikrobielle celler tilpasser sig specifikke stress- og vækstbetingelser og giver konkrete platformstammer og et sæt mutationer, der kan bruges som tekniske værktøjer til at generere produktionsstammer til biomaterialer baseret på vedvarende råmaterialer.

Acknowledgements

I still remember the first day when I started as a Ph.D. student at the Novo Nordisk Foundation Center for Biosustainability (CFB) in the ALE team four years ago. Back then I was excited (and still) about interacting and working with robotics platforms for adaptive laboratory evolution. I had a lot of excitement and rewarding moments, and now the journey has come to an end.

Thanks to everyone helped me, in a way or another, during my Ph.D. studies. In particular to:
My supervisor Adam Feist, I would like to express my sincere gratitude to you for giving me the opportunity to do a Ph.D. at CFB, and who has helped me with lots of guidance, advice and feedback along the way.

Markus Herrgård; it was my pleasure having you as my co-supervisor. I enjoyed working with you. Thanks for your help, support and discussing work and granting me a chair at the SIM group.
I would like also to thank Rebecca Lennen for co-supervising me for two years until she left the CFB to US. Thank you for the help and the support in the lab and providing me with genetic manipulation tools and very valuable protocols.

ALE group colleagues; comrades in arms, Mohammad Radi, Panagiotis, Josfine, Hao and Line thank you guys for sharing the good time and encouragement.

Egyptian friends in Sweden and Denmark; thank you for the nice time we have spent together. Thanks to M. Radi as a colleague and friend for your help and support running my ALE experiment and covering my absence when I was away on Ph.D. classes. I would like also to thank Wael, Abdelrazek and Mahmoud for the nice time we've spend together in Lund. Special thanks to Mahmoud said for reading through the thesis and giving valuable feedback.

Former and current CFB colleagues, thank you for sharing the smile and the good chatting. Especial thanks to Rebeca Thostrup who has guided me through the administrative work till to the end.

Finally, I would like to thank my beloved parents and my small family, for their big love, especially my wife Maryam for your support and understanding during the last four years.

Content

Thesis Organization

The aim of the present thesis is to generate platform strains for raw materials utilization. Adaptive laboratory evolution approach and different techniques were used, such as whole genome re-sequencing, genetic engineering, to achieve the goals. The thesis is divided into three sections, Chapter 1, Chapter 2 then the work that was done presented in three papers with full detailed of generated results, two of which are published and the third is under preparation.

Chapter 1: introduction to microbial cell factories' role in the transformation to bio-based economy

Chapter 2: Conclusion and future perspectives

Paper I: Generation of an *E. coli* platform strain for improved sucrose utilization using adaptive laboratory evolution

Paper II: Generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution

Paper III: Generation of bacterial platform for aromatic acid tolerance and utilization

Chapter 1: Introduction

Summary

This chapter is an introduction to metabolic engineering for developing microbial cell factories and solutions for bio-sustainability and is divided into four sections. The first section gives an initial overview of the chemical industry, the second describes biotechnology and provides a historical context for metabolic engineering, and the third describes evolutionary engineering for metabolic perturbation and how it has been used to accelerate microbial strain design. The last section describes the drive for the generation of platform strains in bio-processing.

Overview of the chemical industry

The industrial revolution marked a major turning point in mankind's history and impacted nearly all aspects of life. One major outcome was the establishment of mass production using machines powered by steam engines, which were subsequently replaced by more energy-efficient internal combustion engines. This led to a structural change in the market from a dependence on renewable resources, such as agriculture for fulfilling the needs for food, clothes, and other products, to industrial-based production. Consequently, nonrenewable fossil resources, such as coal, petroleum, and gas, have become the major sources of both energy generation and chemical and materials production, which has shaped the transformation of our modern society to a crude oil–based economy [1,2].

The chemicals industry represents a significant part of global industry, and its products are of great significance for the prosperity of modern societies. The chemicals industry is estimated to contribute around USD 5.7 trillion to global gross domestic product and to support 120 million jobs (International Council of Chemical Associations, 2019). Chemicals can be found in many product categories, including textiles, safe food supply transportation, housing, health, and hygiene. End-user products include soaps, detergents, tires, clothes, refrigerants, anti-freeze, paints, cosmetics, and pharmaceuticals. Fossil resources are the origin of around 90% of chemicals, particularly organic chemicals, and the remainder is accounted for by chemicals produced from renewable raw materials [3,4].

Fluctuating oil prices and rising concerns about the negative impacts of fossil-based production on the environment have led to growing efforts to find alternative energy and chemical resources. These resources should be cheap and readily available and should not interfere with the food chain or compete with ecosystems. While alternative sources for renewable energy are available, such as solar, hydro-, and wind power, biomass is the only alternative to fossil resources for biofuel energy, chemicals, and materials. Types of biomass include crops, wood, residuals from agriculture, and garbage.

The transformation from fossil resources to biomass requires the development of new technologies to produce biofuels, chemicals, and other materials. In industry,

petrochemical refineries fractionate crude oil into a wide range of products and building blocks for chemical and energy products; similarly, microorganisms can be engineered to utilize biomass and convert it into either fuels or a variety of different biochemical and chemical intermediates that can be further converted to a range of product lines [5–8].

Biotechnology

Biotechnology is based on biology and harnesses cellular and biochemical processes, such as microbial cells and enzymes, for bioconversion of raw materials to target products in processes like fermentation (UN Convention on Biological Diversity, Art 2). Microbes have diverse and versatile metabolisms that allow them to colonize every ecosystem niche on the planet. Throughout history, mankind has used microbes, initially for food and later for medicine, fuel, chemicals, and other biomaterials. The first recorded use of microbes was in the 19th century BCE, by the Sumerians to produce beer and by the ancient Egyptians to prepare leavened bread using yeast [9].

Several products consumed daily are produced using industrial biotechnology on a large scale, including dairy products, such as yogurt and cheese, and pharmaceuticals, such as insulin [10,11]. Microbes can feed on different substrates available in their surroundings to maintain life within the cell. Inside microbial cells, a wide range of chemical reactions occur to utilize a given substrate, either spontaneously or using enzymes as a biological catalyst. Enzymes are responsible for activating a cascade of dozens of biological conversions and chemical reactions inside the cell, which is called metabolism [12]. Microbial cells have the capacity to uptake a diverse range of substrates and are capable of bio-converting them into energy and the building blocks necessary to sustain life within cells or to take part in cell division. There are numerous applications of biotechnology, including animals, plants, agriculture, the environment, fuel, aquatics, and industry [13]

Industrial biotechnology can be used for the bio-production of chemicals and the energy exploitation of microbial cell factories. Wild-type microorganisms (i.e., the original phenotype) have, in many cases, limitations that make them incapable of producing the desired products in high concentrations, yet using biological systems for chemical

production has several advantages over classical chemical production from petroleum refineries, such as stereo-selection of the target chemical [14].

The first attempts to overcome these limitations and disadvantages were through bio-process optimization, including media and fermentation modes or classical strain improvement. One approach to classical strain design was applying mutagenesis to targeted strains, followed by random screening of a large number of isolates. This technique was successful in improving production of, for example, penicillin , but it was laborious and non-reproducible, the metabolic production pathways of the product were not analyzed, and the relevant pathway enzymes were unknown [15]. Bio-process optimization has a number of other limitations, such as product or substrate toxicity that can limit productivity and product yield, and substrate toxicity can arise from the use of raw materials that contain inhibitors that arrest microbial growth and hinder substrate consumption. Strategies are therefore required to generate organisms with optimized cellular metabolisms for industrial use. The success of this avenue requires the development of effective conversion technologies [16].

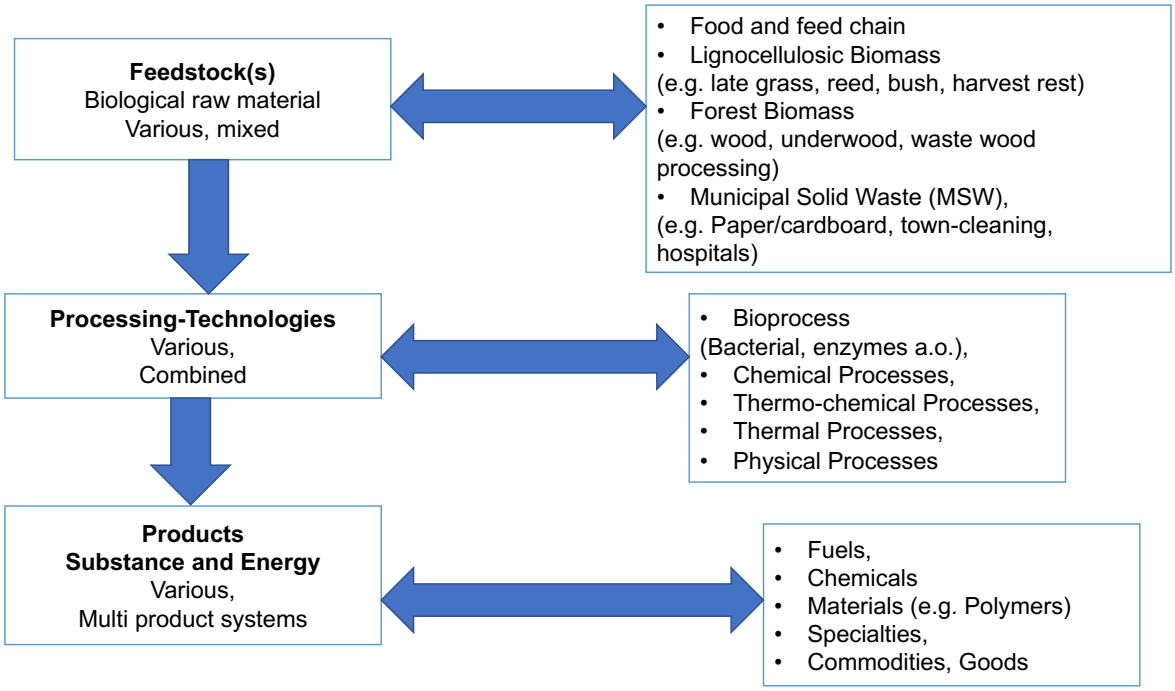


Figure 1: Basic principle of a biorefinery (generation III biorefinery) [17]

There are different routes for converting biomass into energy or industrial products, involving various biological, chemical, and thermal processes. For a particular raw material, a single route can be adopted or a combination of processes. Utilization of raw materials for biochemical conversion in biorefineries may require pretreatment when biomass accessibility to biological systems is hindered by physical and/or chemical barriers. Pretreatment in this context aims to disrupt the close inter-component association between the main constituents of the plant cell wall, to further expose the inter-components for biological conversion using enzymes or biological systems. Biological conversion of raw feedstocks includes two routes: aerobic cultivation and anaerobic digestion.

Biorefineries and Biomass

Significant efforts are being made to shift from a fossil-based economy to a bio-based sustainable economy. Biomass can be used in biorefineries to generate chemicals, energy, and materials, replacing the need for petroleum, coal, and natural gas. Biomass includes anything that can grow and is available in non-fossilized form, including arable crops, trees, bushes, animal by-products, human and animal waste, waste food, and any other waste stream that rots quickly [18]. Biomass is renewable, as plants synthesize chemicals during growth by harvesting energy from the sun and carbon dioxide and water from the environment, releasing oxygen in return, and they can be replaced in years or decades. Developing microbial cell factories for the utilization of biomass can play a key role in ensuring the carbon-neutral cycle of biomass production and consumption and can help to produce energy and chemicals for a more sustainable economy.

Biomass is naturally abundant and can be cheaply planted, strengthening its role in replacing fossil fuel–based production with more renewable resources. The production of plant biomass has been estimated at 170 billion tons annually, only 3.5% of which is used for human needs, which are mainly food and energy generation [19,20]. Biomass in general has a complex composition, similar to petroleum. Plant biomass consists of 75%

carbohydrate, mainly as cellulose, starch, and sucrose; 20% lignin; and 5% protein, lipids, fats, and a variety of other substances, such as vitamins, dyes, flavors, and aromatic essences with very different chemical structures [19]. The main focus should therefore be on efficient access to carbohydrates such as glucose and starch.

Glucose from starch, sugar, or cellulose is accessible by microbial cells through chemical and/or enzymatic hydrolysis of lignocellulosic biomass and is considered a basic broad platform for a wide range of biotechnological products in biorefineries. In this thesis, the term *lignocellulosic biomass* refers to dry plant matter composed of carbohydrate polymers (40% cellulose, 26% hemicellulose), an aromatic polymer of lignin (30%), and other polysaccharides [21,22].

A biorefinery can utilize different feedstocks, integrate different or combined processes, and result in different end products. The basic principles of a biorefinery are shown in Figure 1 [17]. Different factors govern which theme should be adopted for a biorefinery, including demand for the final product and market prices [19]. However, sustainability and a competitive shift to a more bio-based economy require the utilization of biomass residues and wastes of little or no value for the development of microbial cell factories and the production of value-added products. As such, competition for food and cultivation land should be avoided [18,23].

Sugar- or starch-rich feedstocks can be converted using biorefineries into bioethanol as a fuel or other favorable end products (**Paper I**). Lignocellulosic feedstock materials require pretreatment to expose the inter-cellulose to enzymatic hydrolysis, whether through physical processes, chemical processes, a combination of both, or biological processes [22]. Lignocellulose chemical composition can vary from one source of the raw material to another. Several pretreatment methods are available, including acid-based, hydrothermal, and alkaline processing, but most such methods involve side reactions and usually generate by-products that are inhibitory for biological systems, including microbial cells and enzymes. Strategies to overcome the problem of lignocellulosic-derived inhibitors that are formed during pretreatment include selection of microorganisms, evolutionary engineering, and metabolic engineering [22]. This thesis focuses on two remedies for alleviating the inhibition problems: the use of alternative pretreatment solvents, such as ionic liquids (IL; **Paper II**), and an adaptive laboratory evolution (ALE)

approach to generate strains that are tolerant to increasing concentrations of biomass breakdown substrates, such as *p*-coumaric and ferulic acids, and that have improved fitness on glucose (**Paper III**).

Treatment with ILs is one of the most effective approaches for directly releasing monomeric sugars and is an alternative to the pretreatment of lignocellulosic materials [24]. ILs disrupt the non-covalent interactions between lignocellulose components without leading to significant degradation. Cellulose regenerated from IL solutions increases enzymatic convertibility, but, despite its efficacy, any ILs remaining in the pretreated materials are potentially toxic to enzymes and fermentative microorganisms [25]. In paper II, we address this problem through the utilization of ALE to generate platform strains ready to use for deconstructed lignocellulosic biomass.

Metabolic Engineering and the impact of adaptive laboratory evolution

Advances in genetic engineering and recombinant DNA technologies have made rational strain design using precise cell engineering possible. Metabolic engineering is about engineering microbial cell factories for biological manufacturing of chemicals and pharmaceutical products by altering the internal cell metabolism and regulatory processes [26]. The microbial diversity in nature provides a large pool of metabolic networks to choose from, from which a wide range of molecules can be produced in organisms using so-called heterologous production or from which native metabolisms can be improved by applying metabolic engineering. The general schema of metabolic engineering is synthetic biology, protein engineering, and pathway engineering [26]. Synthetic biology deals with DNA and mRNA optimization; protein engineering deals with protein production, enzyme rational design, and directed evolution; and pathway engineering deals with metabolites and fluxes within the cell [26]. Thus, metabolic engineering provides versatile tools for improving microbial cell factories for bio-based production of chemicals, fuels, and materials.

However, using metabolic engineering to generate new microbial cell factories for the production of relevant products at an industrial scale is challenging and expensive. As microbial cells inherently have a complex metabolic network with extensive and complex regulation, any perturbation of specific metabolic pathways for new or native production will have an impact on the regulation of other metabolic pathways. This might be observed as a decrease in growth rate, change in protein concentration, or decrease in product concentration [27]. From six to eight years and over USD 5 million are currently required to develop a new cell factory with a commercially relevant titer [16]. Thus, for more efficient engineering, microbial cell factories are used, involving multiple steps of iteration that consist of four main stages—design, build, test, and learn—in which metabolic design of the target product is implemented and improved through cycles of selection and screening and the data generated at each iteration cycle is used to improve the microbial cell factory design (Figure 2). The cycle is divided into two parts: in-silico work (design and learn phases) and wet-lab work (build and test phases) [16]. Notably, development of a microbial cell factory can start at any point of the cycle.

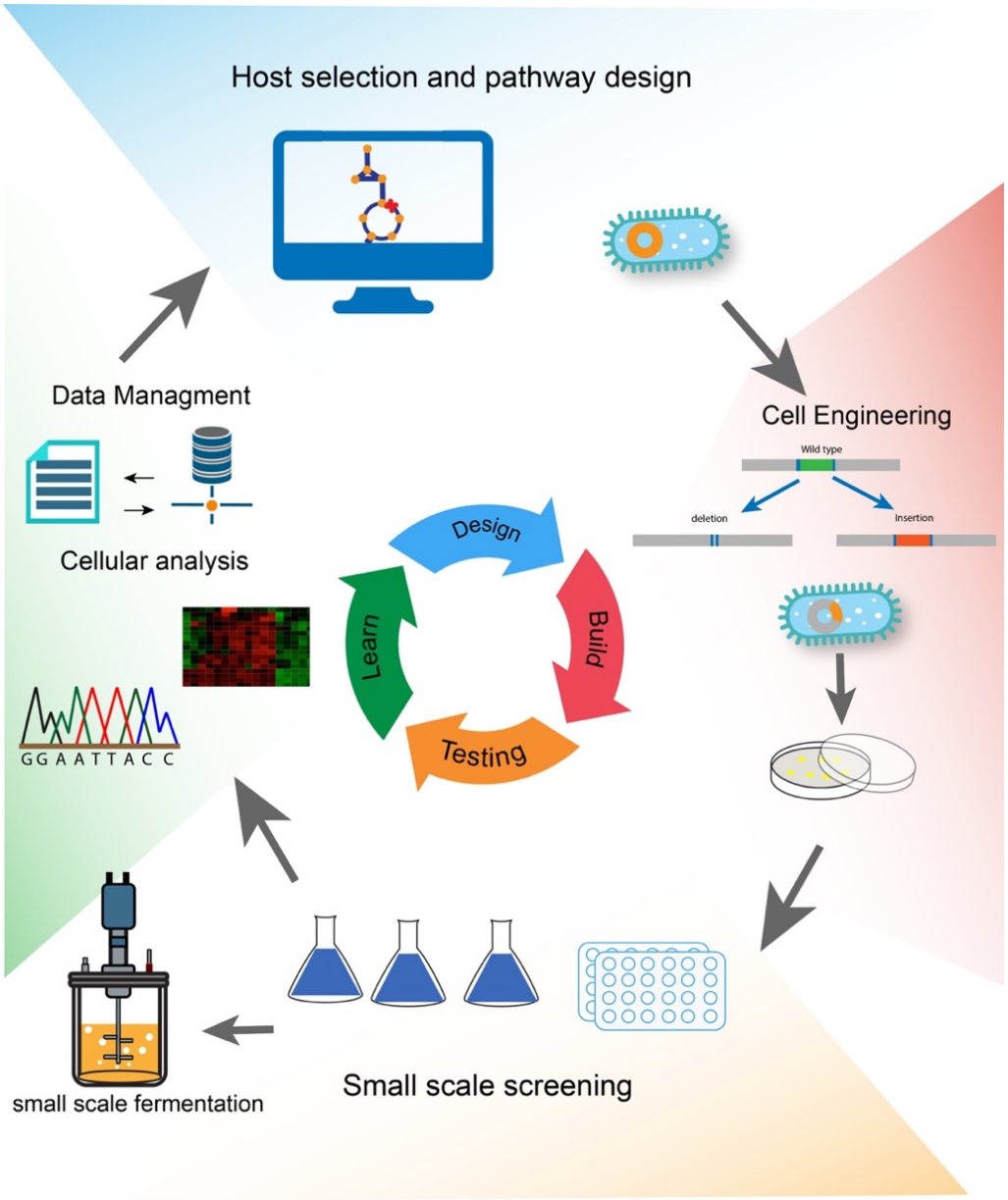


Figure 2: The application of the microbial cell factory cycle to designing cell factories, based on a systematic approach, to generate production strains for heterologous or homologous biomaterials and chemicals.

Alternatively, evolutionary engineering can be employed to improve microbial strains for the production of a desired product or other relevant industrial phenotypes, such as tolerance and utilization of new substrates (Figure 3). To achieve the desired phenotype using evolutionary engineering, the generation of large genetic variations is required, followed by selection and screening [28]. Genetic variations can be achieved by classical strain design methods, such as exposing cells to radiation (e.g., UV or X-ray

mutagenesis) or through chemical mutagenesis, or by applying ALE and in vivo recombination through mating or conjugation [28]. Significant screening or selection processes are then required to identify mutants with the desired phenotype.

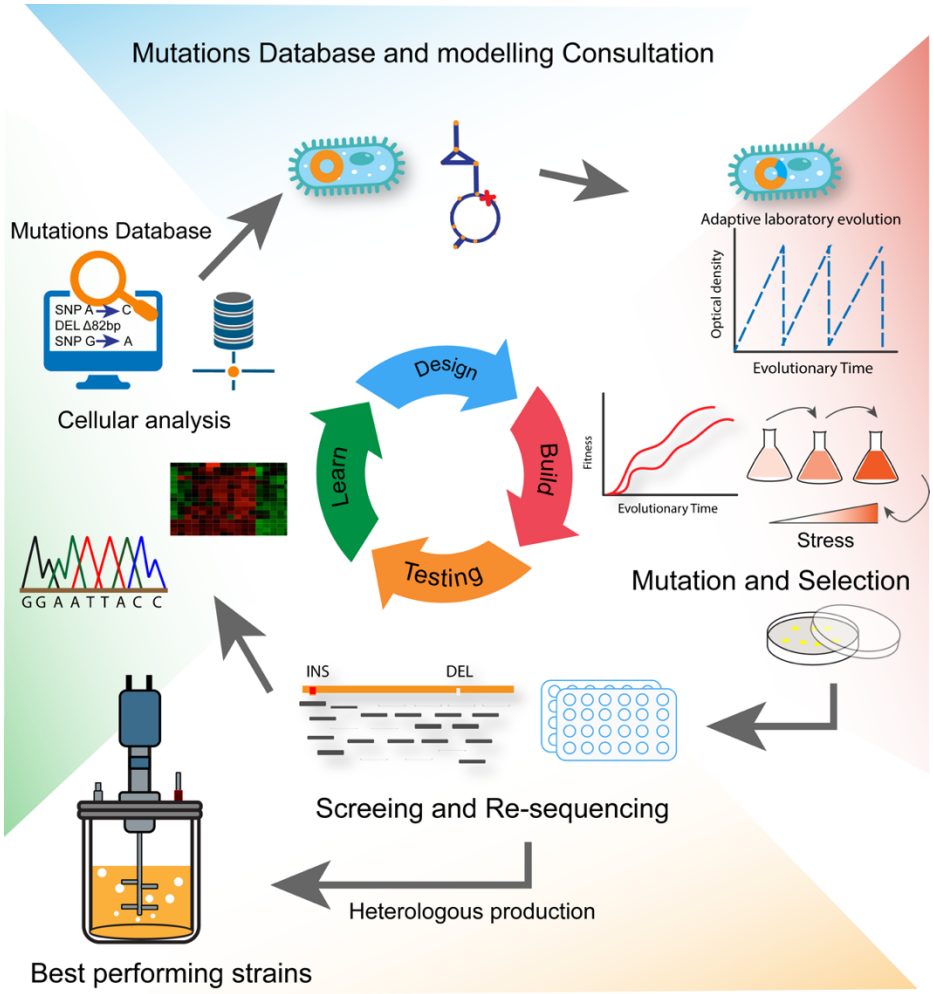


Figure 3: Alternative microbial cell factories cycle of design, build, test, and learn using evolutionary engineering. ALE can be included in the build step for developing strains with the desired phenotypes using sufficient selection pressures to either improve fitness or tolerance. ALE can also be used to replace the design and build steps, in which strains with promising phenotypes can be identified in the test step with screening and re-sequencing. Mutations from different ALE experiments can then be sorted in a database collection for further analysis and used for microbial cell engineering. In the learn step, genotype data is correlated with phenotype data to determine causality and to understand the impact of each mutation on growth.

Figure 4 offers more insight into the role that ALE can play in evolutionary engineering for microbial cell factory development.

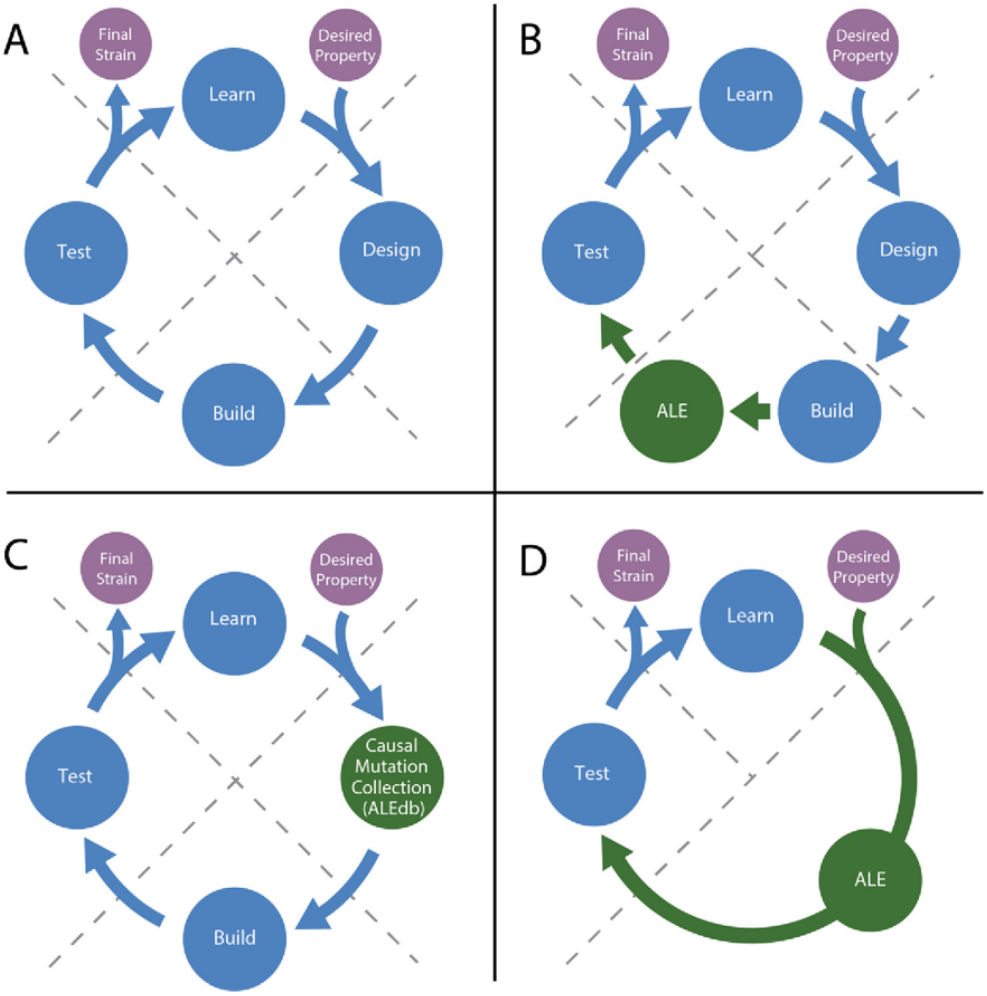


Figure 4: ALE for use in the microbial cell factory cycle, including, design, build, test, and learn. A) The typical design, build, test, and learn cycle used in metabolic engineering to generate a strain with a desired property. B) Augmentation of the cycle in which ALE is included in the build step to improve a strain that has decreased fitness due to a perturbation or to optimize a strain after removal or addition of genetic content. C) Augmentation of the cycle in which a collection of mutations (e.g., ALEdb [29]) associated with a particular phenotype is leveraged for the design step. D) Augmentation of the cycle in which ALE can be used to completely replace the design and build steps and a desirable strain is acquired directly from ALE when a phenotype can be tied to selection without engineering. Figure and most of the legend are adopted from Sandberg et al. (2019) [30].

Design phase

The main purpose of the design phase is planning the required alterations of the host cell. This includes modeling, vector design, and similar in-silico work. The aim of this phase is to identify all possible strategies that can be used to improve the production of a given product. It starts with identifying a good production strain, based on knowledge of the process pH and temperature; the raw material to be used as a growth substrate; strain physiology, such as growth rate, product and substrate tolerance; and the available genetic engineering tools. Once the potential host is selected, identification is made of the pathways required for production in order to achieve a proof-of-concept design. The target product can be achieved by applying the right modifications to the native pathways or by borrowing heterologous pathways and other engineering parts from the diverse microbial genetic pool. The application of metabolic, regulatory, and stoichiometric models is particularly useful in the process design. Applying these models in-silico for the selected host strain before the wet lab is useful for better identifying the required genetic manipulation.

Alternatively, other strategies can be planned in the design phase to achieve the desired phenotype. For example, an ALE experiment can be designed based on a required phenotype that includes increased tolerance to a given chemical or toxic substrate, improved fitness, or improved growth using a new carbon source [30,31].

Build phase

This phase is also called the construction or synthesis phase. The build phase covers any type of genetic engineering and includes screening and selection, synthetic biology, evolutionary engineering, genetic engineering, and metabolic engineering. Modification planned in the design phase is introduced to the host using genetic engineering tools. Several approaches can be implemented using rational design to ease toxic stress by applying modifications to the metabolic pathways. A non-rational approach can also be used, such as the ALE method for the selection of improved phenotypes under the constant pressure of selection [8,32].

Applying evolutionary engineering, the build phase can be replaced by a mutation and selection phase (Figure 3). This phase consists of introducing mutations into the host strain and screening for the best performing strains. Means of introducing genetic variations include classical strain design with radiation or ALE.

ALE can provide more sophisticated phenotype selection through introduction of genetic variations under selection pressure to generate potential host strains with improved phenotypes. When a desired phenotype can be linked to growth, such as production of a chemical or utilization of a substrate, then selection of the improved phenotype can be achieved by selecting the fast-growing strain [33].

Test phase

This is also called the characterization phase. The test phase consists of any type of cultivation and/or quantitative physiology, including bioreactor growth experiments and physiology studies, such as growth rates, yields, products, concentrations, and morphology. In this phase, the performance of the built strain is evaluated in terms of tolerance, expression levels, and different genetic engineering modifications that result in the production of the desired chemicals [16]. Typically, a large number of clones are tested, which requires implementation of high-throughput screening. Target molecules and their concentrations can be detected using analytical assays involving high-pressure liquid chromatography, with or without other assays, such as mass spectrometry or gas chromatography (GC) [34].

If evolutionary engineering is applied to select a phenotype through mutation and selection, it is necessary to characterize the mutations behind that phenotype. Mutation identification can be achieved using whole-genome re-sequencing techniques. Advances in next-generation sequencing in the last decade and the low cost of re-sequencing each nucleotide have revolutionized rapid whole-genome re-sequencing. Identified mutations can now be listed to determine the causality behind a unique phenotype, such as growth behavior or production of a desired metabolite. Mutants expressing the desired phenotype can also be characterized in terms of production using any available quantitative method,

such as liquid chromatography or other high-throughput assays relevant to the expressed metabolite.

Learn phase

This is also called the analysis phase. The learn phase includes an analysis of all types of additional data sets generated. Examples of data generated include omics-data, such as genomics, transcriptomics, proteomics, population heterogeneity, and metabolic flux measurements. Generating production strains by applying classical strain design or using ALE can also lead to accumulation of “hitch-hiker” mutations along with the causal mutations that are behind the desired phenotype. Several approaches have been investigated to solve this problem by linking the unique phenotype to the genotype. For example, causal mutations can be easily identified by running multiple parallel replicated ALEs and re-sequencing both the populations and the isolates derived from them to find key genes or genetic regions that have mutated several times or to find the same mutation occurring across independent populations [35,36]. The effect of each new mutation on protein activity can also be sought using genome modeling [37].

In order to select a promising microbial cell factory host strain for the production of target chemicals or materials, there are certain criteria to consider for the successful generation of a production strain [38].

Characteristics of a good microbial cell factory

When selecting the host strain for metabolic engineering, the strain should be culturable—with reproducible rapid growth and known required nutrients—and genetically engineerable—with the capacity to express a gene/protein of interest. Another factor to consider is the availability of previous experience controlling the growth and engineering the host cells of the selected host strains. Moreover, the host should be GRAS (generally regarded as safe), and the production process should be approved with no toxic byproduct formation and with assurance that the product has a potential market and can be sold [39]. Another key to successful cell factory is having a match between the product and the host organism where the product is considered safe for the cells and avoids the production of conditions that are hostile to the product. Preferably, natural producers of the desired compound are a good choice for production host cell factories. Alternatively, the production process should be scalable from milliliters (on a lab scale) to millions of liters (on an industrial scale). A process that only works in a lab scale is hard to make financially viable for cheap products on an industrial scale. Several factors can serve as indicators for the possibility to scale up the production process: the product yield, the product concentration, and the volumetric productivity and the ease of purification. However, the final decision is always controlled by the production process economics [40]. Thus, the quest for cheap substrates is one of the major goals; the price of substrate and product must be balanced. This balance can be achieved with the utilization of renewable substrates.

Adaptive laboratory evolution

Microorganisms can survive and adapt to different environments via the acquisition of beneficial mutations by random genomic mutations and subsequent selection. Spontaneous mutations occur naturally in microorganisms and can lead to changes on the genome level, affecting their chance of survival. Therefore, microbial cells can adapt to different environments by changing and selecting for metabolic and regulatory network perturbations for optimal growth—which is called adaptive evolution [35,41]. Interestingly, application of ALE can range from performing basic science to developing a microbial cell factory for industrial biotechnology applications. Applications of ALE are widespread and include the activation of latent pathways, growth rate optimization, environmental adaptation, general discovery, substrate utilization, increased tolerance, increased product yield, and titer [30].

With the application of the right conditions in the lab, metabolic perturbation can be directed for various applications. During the ALE experiment, microorganisms are cultivated for prolonged periods of time, which can range between weeks to several years and can allow for improved phenotype selection. Microbial cells are well suited for ALE as they grow fast, have simple nutrition requirements, and maintain large populations [42].

There are several approaches to performing ALE; however, here presents three of the general methods for performing adaptive laboratory evolution on a lab scale: serial transfer, chemostat culturing, and colony transfer. (see Figure 5 below). The first approach to performing ALE is batch cultivation of serial transfers in shake flasks. Shake flasks can be replaced with deepwell plates for higher throughput. Cells are propagated in parallel serial cultures while small culture aliquot are transferred to a new flask, with a fresh medium, for another round of growth. This approach is characterized by its simple and cheap setup to operate and its capability for multiplexing (i.e. running large numbers of parallel cultures). In this context, several growth conditions, such as temperature or chemical stress, can be controlled to obtain evolved microbes for a selected phenotype.

However, batch cultivation has some downside factors which include varying biomass, fluctuating growth, and other fluctuating environmental conditions such as pH and dissolved oxygen from one batch to another [43]. However, consequences of these factors on the experimental setup may not be important.

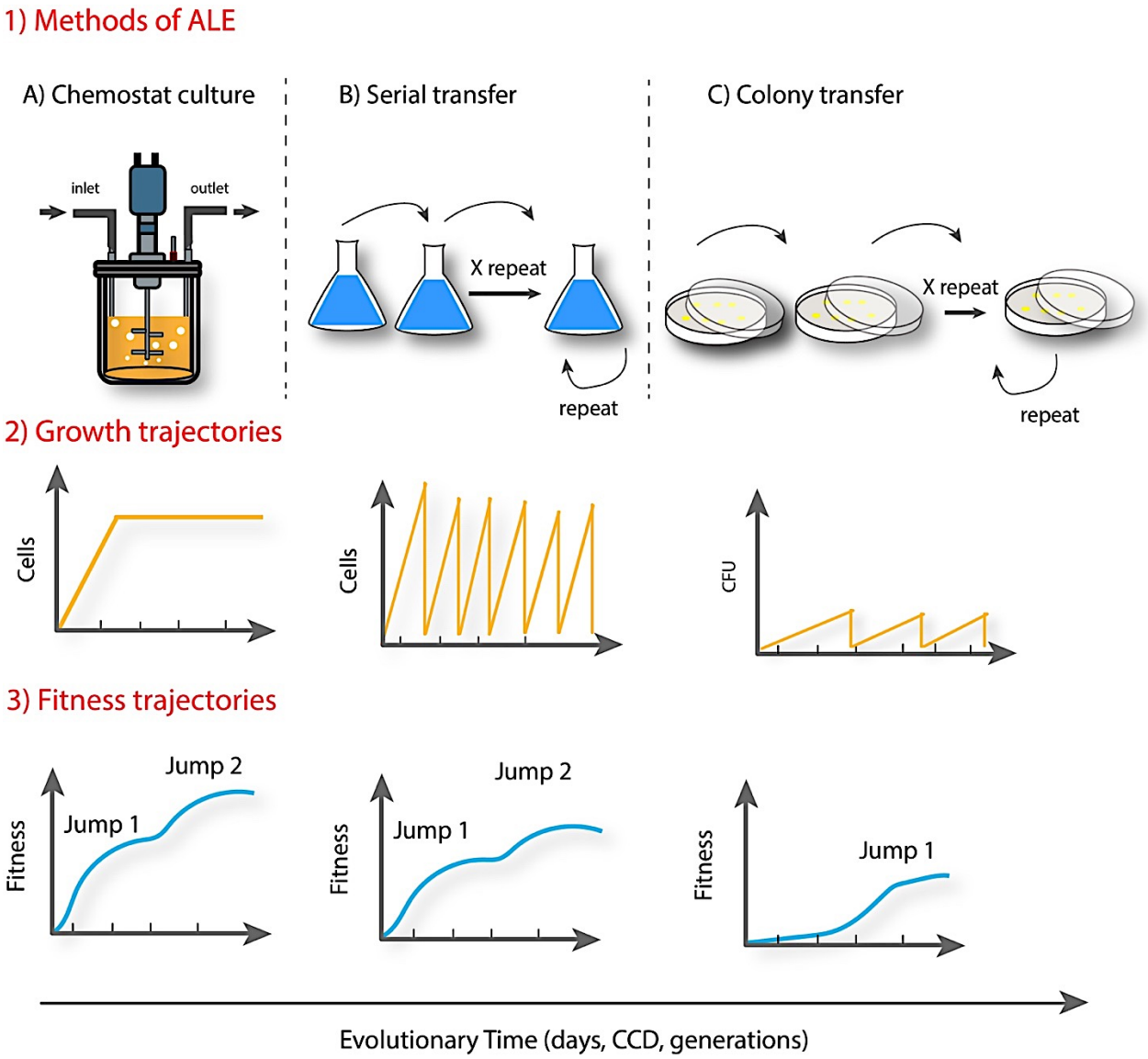


Figure 5: Methods of adaptive laboratory evolution: A) Chemostat, B) Serial transfer, C) Colony transfer and their respective illustration of changes in cell growths and fitness trajectories.

The second approach to performing ALE applies chemostat or continuous culture in bioreactor vessels. Cells are cultured for a steady state in a bioreactor in which fresh

medium is continuously added, while culture liquid containing cells and leftover medium are continuously removed. Furthermore, environmental factors such as nutrient supply, pH, and oxygenation can be tightly controlled. However, chemostat culture still requires complicated methods to operate and can have major drawbacks compared to serial transfer ALE when running parallel vessels. Another major difference between serial transfer and chemostat culture ALE while selecting for growth phenotype, is the presence of nutrient-sufficient or nutrient-limited conditions, respectively. Specifically, in order to control chemostat cultures, growth rates below the maximum growth are applied. In contrast, serial transfer cells are transferred before the stationary phase of growth during the exponential phase of growth with surplus nutrients. Therefore, cells are kept under exponential growth in serial transfers—rather than kept constant in chemostat—by limiting major growth nutrients such as glucose and nitrogen sources [43]. In this context, batch serial transfer ALE has advantages over chemostat when selection is required for higher fitness (i.e. growth rates or reliance on biomass yields rather than a yes-or-no survival phenotype as in the chemostat culture).

Finally, the third common method for performing ALE is colony transfer. In this method, cells are allowed to grow on solid agar plates in the selected environment. Colony transfer ALE can be useful for phenotype selection that is based on yes-or-no survival phenotypes such as antibiotic resistance studies in which cells are exposed to high antibiotic concentration and growth is permitted for multiple days [44]. However, the number of replications and the population size are very small when growing cells on agar plates, compared to the serial transfer and chemostat methods. This can eventually lead to decreased variation in the cellular state and may result in a prolonged ALE experiment time to accumulate the required mutation for a required phenotype. The emergence of whole genome re-sequencing, and the drop-in cost of re-sequencing, combined with ALE allowed for observing mutational dynamics and the diversity of evolved populations. Thus, ALE can be of great utility in the microbial cell factories cycle used in strain development, enhancing or even replacing the design and build steps.

Applications of adaptive laboratory evolution (ALE)

As previously mentioned, the applications of ALE are widespread but can be summarized into five main categories [30]: growth rate optimization, general discovery, substrate utilization, increased tolerance, increased product yield, and titer. Each of these categories will be addressed briefly in the following part to highlight their main applications. Several review publications for further comprehensive summary on ALE applications and success stories was published by Sandberg et al., 2019 [30].

a) Growth rate optimization

One of the key factors in selecting host strains for viable bio-production process is how fast they can grow or having short generation times. We can apply ALE to optimize the growth rate of industrially relevant microbial species or improve growth defects in engineered strains or even answer a question about fundamentals behind growth rate change in a specific environment or perturbation [45,46]. Several studies have successfully used ALE in this quest; [35,47,48]. In this thesis, we have successfully applied ALE (Paper I) to fix fitness defects in rationally designed strains in order to restore microbes back to their wild physiology and even push it further [49,50].

b) General discovery

ALE can be as well applied in strain design and optimization through analyzing the evolutionary outcomes in a controlled laboratory setup. Using the knowledge acquired from responses to genetic perturbations and sub-optimal growth environment [30] can significantly help in the strain design cycle. Moreover, ALE can be used to study metabolic networks and regulations and can further paired with metabolic modeling approaches to conduct systems-level analysis.

c) Substrate utilization

The need to develop economically visible bio-production processes requires lowering the cost of cultivation substrate needed for production strains. This can be done by using renewable raw resources which can be sometimes inaccessible to biology systems because of lacking the ability to utilize them or the toxicity of the raw materials. For this purpose, we can design ALE experiments to enable improved or de novo substrate uptake and utilization. Improving substrate utilization was explored in Paper I and Paper III.

d) Increased tolerance

Genetically engineered strains to produce metabolite of interest in high titer usually leads to microbial growth inhibition or using raw substrates such as lignocellulosic biomass that usually contain cellular stressors for bacterial growth. ALE has proven to be a valuable tool to overcome these stressors such as sub-optimal pH levels [50], osmotic stress [43], temperature [51], UV irradiation [52], nutritional stressors [53]and byproducts of biomass pretreatments [54]. Improving substrate and raw materials deconstruction solvent tolerance were explored and investigated in paper 3 and 2, respectively.

e) Increased product yield, and titer.

Eventually, the main goal of any microbial cell factory is to maximize the final product concentration for a given fermentation process. However, maximizing production of interest metabolite using the traditional rational strain design is frequently insufficient and requires intensive prior knowledge about strain metabolism. Metabolite of interest can be coupled with for example to biomass formation in a growth coupled manner then apply ALE to improve the growth rate [33].

Motivation for the Generation of Platform Strains in Bioprocessing

As explored above, it is technically already possible to produce biomaterials, including chemicals and biofuel, from renewable resources by using microbial bio-refineries based on cell factories such as bacteria and yeast. Microbial cell factories can substitute mankind's complete dependency on fossil fuels, which can have a major positive impact on the environment and initiate the transition from a petroleum-based economy to a sustainable, bio-based economy. The main challenge in developing microbial cell factories, is achieving balance between production cost of input nutrients and the generated product price in order to compete with fossil-derived products; the process has to become much more efficient with innovative approaches. Therefore, there is an obvious need for platform strains that can be readily utilized to make desirable products from sustainable feedstocks. Such strains would rapidly enable the development of consolidated bio-processing strains and significantly cut the capital costs necessary to generate commercially feasible production processes.

Platform strains, which can be readily utilized to make desirable products from sustainable feedstocks, would rapidly enable the fast development of consolidated bio-processing strains. The term platform strain was first coined by Nielsen et al. to describe yeast strain that was ready to produce different natural products [55]. Several advantages of using platform cell factories include their readiness to be used with well-characterized genetics and physiology, the ability to easily attain product approval (as they have already been used as cell factories), and the availability of genome editing and gene expression tools (plasmids, promoters, and terminators). However, developing successful platform strains using classical metabolic engineering is expensive and time-consuming and requires a comprehensive knowledge of the microbial metabolism and regulation process. Therefore, speeding up cell factory development is one of the main barriers.

The emergence of adaptive laboratory evolution (ALE) and whole genome sequencing (WGS) over the past decade has allowed for rapid generation of optimized strains and

the understanding of the genetic basis of observed phenotypes without prior knowledge of how the cells function.

At the Center for Biosustainability (CFB), we have engineered a first-class pipeline combining these two technologies to harness the power of biology and uncover biological "parts" (i.e. key genetic mutations) that can be utilized for applied biotechnology. Over the course of the Ph.D. project, this pipeline was used to develop platform strains and optimized for the utilization and tolerance of raw material feedstocks (such as lignocellulosic hydrolysates and raw sugar sources).

Microbial cell factories based on biomass as renewable resources, such as lignocellulosic materials, can cut down the cost of developing an economically feasible production process. However, in order for microbial systems to gain access to the fermentative substrate for microbial utilization, biomass needs to be deconstructed into monomeric building blocks. Therefore, a pretreatment process is required to remove the chemical and physical barriers and expose the cellulose for enzymatic degradation to release fermentable sugar. However, pretreatments often involve side reactions and result in lignocellulosic-derived byproducts that are inhibitory to the microbial processes [22]. There are several methods for lignocellulosic pretreatment, but one of the most effective approaches is to release monomeric sugars through the treatment using ionic liquids (IL) [24]. ILs are promising organic solvents and are effective for deconstruction and generation of high sugar turnover from feedstocks—without significant degradation—and producing limited inhibitors [22,56,57]. Despite it is efficacy, the remaining amount of IL after the pretreatment is very toxic to biological systems—including enzymes [25]. Chapter 3, manuscript I explains how the generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution has been achieved.

Similar to lignocellulosic biomass, sugar cane juice, rich with sucrose, can serve as a main carbon source because it is a cheap, abundant, and renewable feedstock very rich in sucrose [58,59]. Sucrose is an attractive industrial carbon source due to its abundance and the fact that it can be cheaply generated from sources such as sugarcane and sugar beet. However, only a few characterized *Escherichia coli* (*E. coli*) strains are able to metabolize sucrose, and those that can are typically slow growing or pathogenic strains.

The slow growth issue was addressed in **manuscript II**, which presents the generation of an *E. coli* platform strain for improved sucrose utilization using an adaptive laboratory evolution.

There are several strategies to counter the inhibition problem that arises when deconstructing lignocellulosic biomass. One of them is to select for microorganisms with a natural resistance to inhibitors. For example, *Pseudomonas putida* strain can serve this role given its natural tolerance toward several chemicals and its ability to metabolize exotic chemicals such as toluene. Pseudomonas putida is a promising production chassis, as it can metabolize aromatic compounds, particularly those derived from lignocellulosic biomass. However, the lignocellulosic biomass degradation process typically yields high titers of aromatic acids at levels harmful for microbial cells. Manuscript III presents how the generation of bacterial platform strains for aromatic acid tolerance using adaptive laboratory evolution was obtained.

The main workflow implemented throughout this thesis in a quest to generate platform strains for three representative conditions of raw material utilization is outlined in three major steps (see Figure 6):

1. Utilize automated adaptive laboratory evolution to obtain optimized bacterial strains that can efficiently tolerate and consume raw materials

The utilization of the automated adaptive laboratory evolution machine (ALE machine) has several advantages. The ALE machines can maintain a large number of independently evolving cultures, continuously tracking their growth rates and keeping them growing around the clock by serially passing large volumes of the cultures (hundreds of millions of cells) to new culture tubes once mid-exponential phase has been reached, thus maximizing the chance of beneficial mutation fixation in the population. The ALE machine allows for continuous tracking of the culture growth rates, and once a specific phenotype is achieved (e.g. high growth rate or tolerance phenotype) then termination of the ALE experiment can be evaluated. This aim will not only yield optimized endpoint strains (Figure 6) but also enable the study of a mutations dynamic relative to a

specific phenotype, as well as a ‘fitness trajectory’ of their steadily increasing growth rate over the course of the experiment. The strict controlled growth and passing routine can also help prevent the intrinsic problem with batch cultivation due to fluctuations in growth and growth environment which can compromise the applied selection pressure.

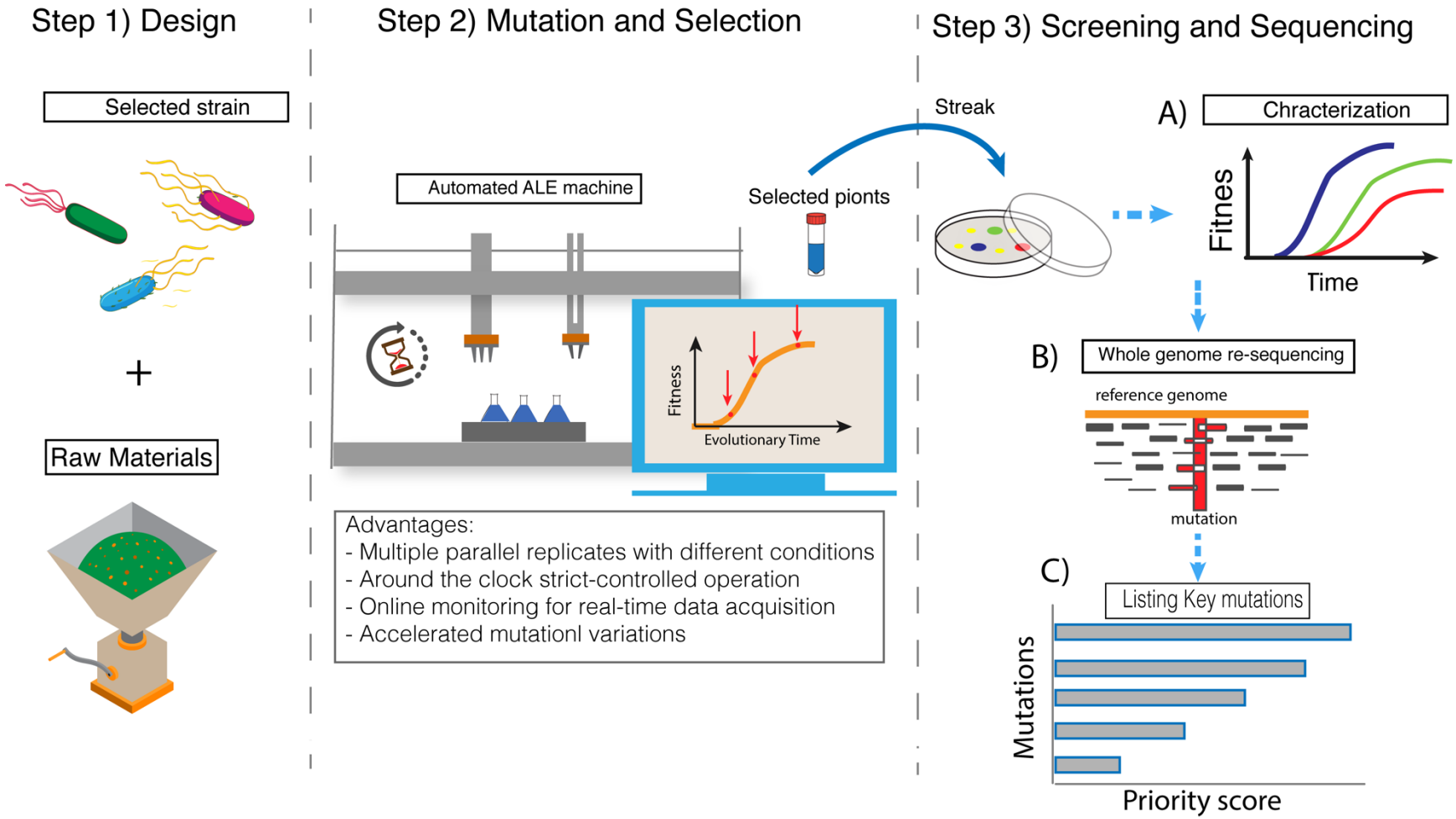

Step 1) Design
Selected strain
+
Raw Materials
Step 2) Mutation and Selection
Automated ALE machine
Selected pionts
Fitness
Evolutionary Time
Advantages:
- Multiple parallel replicates with different conditions
- Around the clock strict-controlled operation
- Online monitoring for real-time data acquisition
- Accelerated mutationl variations
Step 3) Screening and Sequencing
Streak
A)
Chracterization
Fitnes
Time
B)
Whole genome re-sequencing
reference genome
mutation
C)
Listing Key mutations
Mutations
Priority score

Figure 6: Generation of platform strains work flow. The work flow outlined was executed to generate platform strains capable of robust growth on raw materials.

2. Perform whole-genome sequencing on evolved strains to uncover key enabling mutations for consumption and tolerance.

Optimized strains (i.e. clones generated from the first step) were screened for isolates with desired phenotypes. Optimized strains were afterward resequenced using our in-house sequencer and bioinformatics pipeline to determine key enabling mutations [60]. A comparison across the multiple independent replicate ALE experiments was used to generate a prioritized list of the likely key enabling mutations, both strain-specific and potentially general to the species, for the desired phenotypes.

3. Utilize genome-engineering and clonal analysis to generate platform strains optimized for raw material consumption and reintroduce key mutations into a production strain.

The key mutations from step 2, will were reintroduced into clean starting strains, and then they were characterized. On the other hand, mutations occurring in the optimized strains will were reverted to wild-type alleles to evaluate their importance. Promising mutations will eventually be introduced to the existing production strain with a relevant chemical product to test its potentials. Occasionally, re-introducing key mutations were not necessary as there were instances where isolated clones almost carry all identified key mutations.

Using these approaches, the generated platform strains under this project can be marketed as cell factories for bio-based chemical production using renewable feedstocks.

Chapter 2: Conclusion and future perspective

Summary

This chapter contains a conclusion of the work done in this thesis and plans for future work. The first section gives a highlight on the work generated as published papers. The second section describes future work perspectives.

The work performed in this thesis represents a valuable effort to generate platform strains for raw material utilization. Raw materials are a commercially attractive feedstock for bio-production because they are cheap, abundant, and renewable. However, their use is hindered by poor consumption (by most microbial strains) or because they contain mildly toxic compounds that retard growth. Fortunately, ALE was successful in tackling both of these issues. Although ALE could be applied to any raw feedstock, efforts were focused on three main materials: (1) ionic-salt hydrolyzed biomass, (2) raw sugarcane juice, and (3) grass family hydrolysate breakdown products. In short, these materials were identified as biomass sources that can serve as cost-effective routes to sustainable production around the world [61,62]. The selected start strains and raw materials were case studies of the developed workflow for platform strain generation. The developed work flow can be further expanded to include different raw materials and new strains.

The following section contains a summary of the work accomplished for generating platform strains for raw material utilization, especially sugarcane juice, lignocellulosic deconstruction hydrolysate using ionic liquids, and lignin feedstocks, paper I, paper II and paper III, respectively.

As mentioned previously, ALE can also be used to improve strain fitness after genetic perturbation. **Paper I** report the work done to engineer non-native sucrose utilization into additional *E. coli* strains has been investigated, but the generated strains were characterized by slow growth rates on sucrose as compared to glucose. Escherichia coli K-12 MG1655, as a non-native sucrose consumer, was engineered and then evolved using ALE towards higher growth rates on sucrose alongside an *E. coli* W strain as a native consumer control. Evolved K-12 clones displayed an increase in growth and sucrose uptake rates when tested on sugarcane juice, whereas *E. coli* W clones showed an increase in sucrose uptake rates without a significant increase in growth rate. Key mutations identified for the engineered K-12 strain were found in RNA polymerase subunits and metabolic correction mutations, whereas the E. coli W mutations were found in sucrose metabolism regulation. Validation of mutation causality using knock-in strains showed significantly improved fitness on sucrose as well as for glucose. The generated K-12 and W platform strains, and the specific sets of mutations that enable their

phenotypes, are available as valuable tools for sucrose-based industrial bioproduction in the facile E. coli chassis.

Paper II deal with the utilization of a systematic ALE approach to generate the platform strain's tolerance to an ionic liquid called Tolerance Adaptive Laboratory Evolution (TALE). The method was developed and conducted with a process-controlled automated platform. Two *E. coli* strains, K-12 MG1655 and DH1, were evolved to tolerate high levels of Ionic Liquids (ILs) for use as platform strains when utilizing deconstructed biomass. The best-performing strains derived from the TALE approach demonstrated robust growth at 8.5% (w/v) and detectable growth at up to 11.9% (w/v) of the ionic liquid [C2C1Im][OAC]. The generated strain presents, to our knowledge, the best performing strains for elevated ILs concentration. Key mutations in highly tolerant strains that were shared across many parallel replicates were identified in transport processes associated with the functions of *mdtJ*, a multidrug efflux pump, and *yhdP*, an unknown transporter. The best-performing strains represent platform strains tolerant of ILs, amenable for their expression of heterologous pathways and display of superior performance to rationally engineered IL tolerant strains. In fact, the best performing strains from this study. after further testing in a collaboration study, showed improved growth in the presence of four other ILs. Furthermore, these best performing strains showed significantly increased heterologous production of isoprenol compared to the control strain. Interestingly, the derived strain has been shown not only to tolerate ILs but also consume ILs as a carbon source [63].

Paper III present the work done to generate strains capable of using and tolerating high concentrations of aromatic compounds; adaptive laboratory evolution was utilized to improve the growth properties of Pseudomonas putida KT2440 strain in the presence of two common lignocellulosic feedstock breakdown aromatic acids: *p*-coumaric and ferulic acid. The generated platform strains showed improved fitness and tolerance levels over the background strain. Key occurring mutations were knocked out from the background strains, and the resulting strains were characterized. The causality of the mutations related to the gene products of *algE* and *ttgB* were confirmed via gene deletion experiments, and it was found that they conferred a growth benefit when compared to the starting strain.

Several experimental tests have been left for the future due to lack of time; follow up experiments are time consuming and require intensive work and collaborations between different fields to accomplish. For example, future experiments concerning demonstration of the performance of the generated platform strain under a heterologous production setup will be considered for platform strains generated for utilization of sucrose and aromatic acids. This can be done using the generated platform strains and the identified causal mutations to optimize pre-existing production strains for growth on raw materials Figure 1. At present, an ongoing collaboration work is concerned to develop sucrose utilization in L-serine production strains [64] and test production in 10 Liter fermenter at the CFB. The second effort is to use the best performing strains for aromatic acid utilization for muconic acid production as a collaboration work with national renewable energy laboratory, US. This could not be achieved by simply performing ALE starting with a production strain because the energetically expensive product synthesis would be strongly selected against, yielding strains that grew better but had lost their production capabilities.

Furthermore, a follow up experiment is needed to better understand the mechanistic role of key identified mutations, especially the mutation behind the tolerance phenotype for the improved fitness on aromatic acids tolerant strain phenotypes and ionic liquid tolerance phenotype with multiple omics assays to better understand the underlying mechanisms, as performed previously [65]. Revealing the mechanistic role of the key mutations will add to our knowledge about their respective mechanism and help developing the improved phenotype in other production strains.

Step 4) Genetic Engineering validation

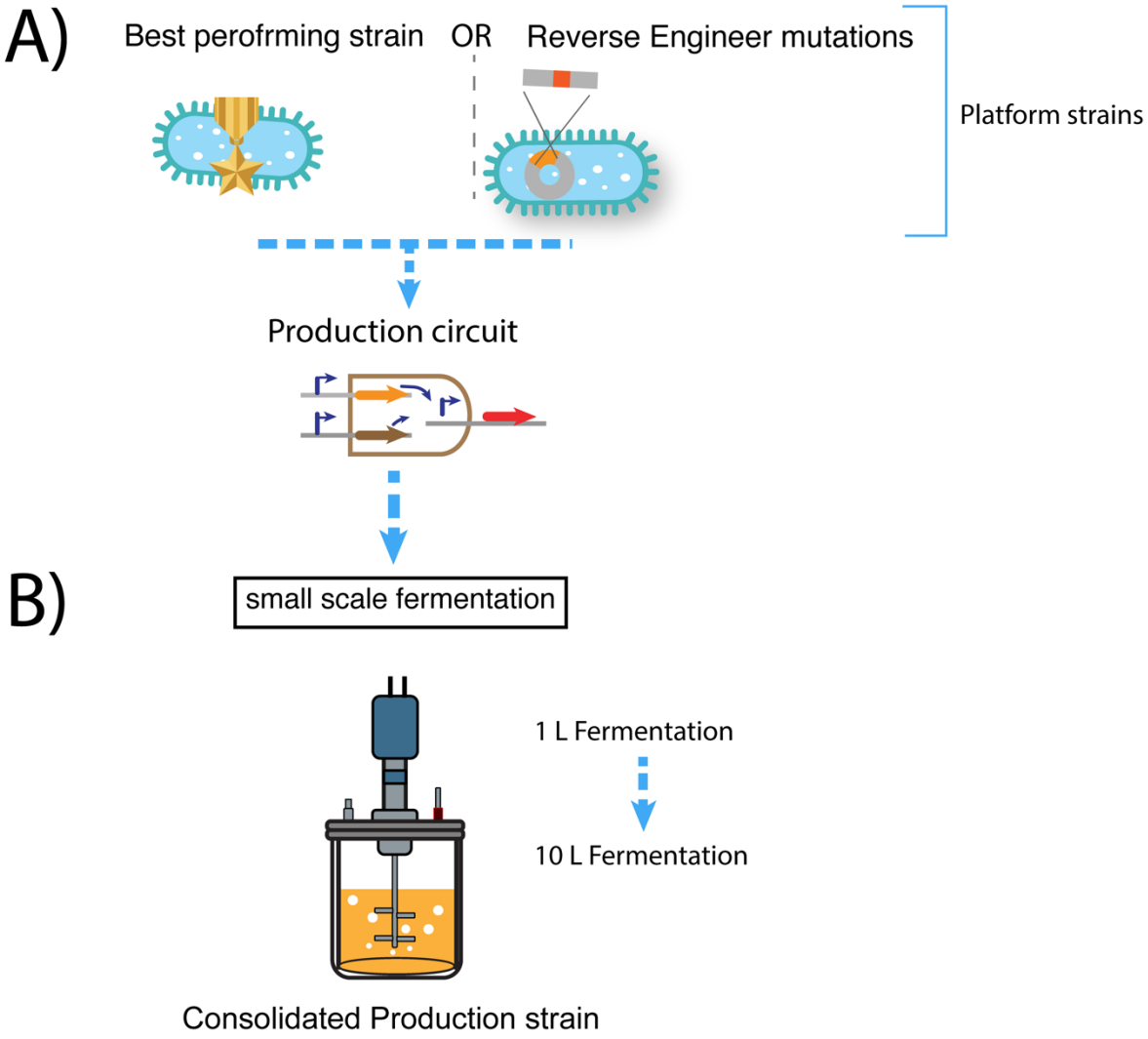


Figure 1: Future follow up work to validate the generated platform strains for production processes. The best performing strains based on the screening steps or the best performing strains with reverse engineered selected key mutations will serve as a production chassis. Small scale fermentation can then be used for production validation.

In summary, this dissertation in its entirety demonstrated a number of main findings through the application of a generalized ALE-based workflow for platform strain generation. These findings showed that ALE has proven to be a valuable tool for effective generation of platform strains given the right selection pressure and environment conditions. Effective generation of different platform strains for three different conditions ready to use for raw material utilization based on lignocellulosic feedstocks and sugar cane juice, suitable for use as production chassis for viable products. Moreover, several genes with hypothetical functions were identified as causal mutations for specific phenotypes. This further reflect the potential use of ALE for general discovery of gene functions.

The developed workflow in this thesis, including performing ALE, tolerance ALE, NGS and further validation of the causal mutations, shows its potential utility so that it can be used in similar research. Although there is future work to do, these results are informative and created tangible strains for bioproduction. They establish ALE as a powerful tool for engineers and biologists to solve the problem of the utilization of raw feedstocks.

References

1. Murmann JP. Chemical Industries after 1850 Oxford Encyclopedia of Economic History [Internet]. 2002. Available from: http://johann-peter.murmann.name/chem-industry.htm.

2. Teich M, Porter R. The industrial revolution in national context : Europe and the USA [Internet]. Cambridge University Press; 1996 [cited 2019 Sep 15]. Available from: https://catalog.loc.gov/vwebv/search?searchCode=LCCN&searchArg=95025377&searchType=1&permalink=y

3. Spiro TG, Stigliani WM. Chemistry of the environment. Prentice Hall; 2003.

4. Vollheim T. Survey of industrial chemistry. ByPhilip J. Chenier, 2nd Revised Edition, VCH, New York 1992, 527 pp., hardback, $ 95, ISBN 1-527-28186-X. Adv Mater [Internet]. John Wiley & Sons, Ltd; 1993 [cited 2019 Sep 15];5:400–400. Available from: http://doi.wiley.com/10.1002/adma.19930050520

5. Kamm B, Kamm M. Biorefineries - Multi product processes [Internet]. Adv. Biochem. Eng. Biotechnol. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006 [cited 2019 Aug 29]. p. 175–204. Available from: http://link.springer.com/10.1007/10_2006_040

6. Fadel M, Keera AA, Mouafi FE, Kahil T. High Level Ethanol from Sugar Cane Molasses by a New Thermotolerant Saccharomyces cerevisiae Strain in Industrial Scale. Biotechnol Res Int [Internet]. Hindawi; 2013 [cited 2019 Jun 4];2013:253286. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24363937

7. Li L, Li K, Wang Y, Chen C, Xu Y, Zhang L, et al. Metabolic engineering of Enterobacter cloacae for high-yield production of enantiopure (2R,3R)-2,3-butanediol from lignocellulose-derived sugars. Metab Eng [Internet]. Elsevier; 2015;28:19–27. Available from: http://dx.doi.org/10.1016/j.ymben.2014.11.010

8. Mundhada H, Miguel JS, Schneider K, Koza A, Christensen HB, Klein T, et al. Increased production of L-serine in Escherichia coli through Adaptive Laboratory Evolution. Metab Eng [Internet]. Elsevier; 2016;39:141–50. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1096717616302348

9. Bhatia S, Goli D. Introduction to Pharmaceutical Biotechnology, Volume 1 Basic techniques and concepts [Internet]. IOP Publishing; 2018 [cited 2019 Aug 26]. Available from: http://iopscience.iop.org/book/978-0-7503-1299-8

10. Coffey AG, Daly C, Fitzgerald G. The impact of biotechnology on the dairy industry. Biotechnol Adv [Internet]. 1994 [cited 2019 Sep 16];12:625–33. Available from: https://linkinghub.elsevier.com/retrieve/pii/0734975094900035
11. Lancini G, Demain AL. Bacterial Pharmaceutical Products. The Prokaryotes [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013 [cited 2019 Sep 16]. p. 257–80. Available from: http://link.springer.com/10.1007/978-3-642-31331-8_28
12. Doelle HW. Bacterial Metabolism. Elsevier Science; 1975.
13. Information Resources Management Association. Biotechnology : concepts, methodologies, tools, and applications.
14. Ren Q, Ruth K, Thöny-Meyer L, Zinn M. Enatiomerically pure hydroxycarboxylic acids: current approaches and future perspectives. Appl Microbiol Biotechnol [Internet]. Springer-Verlag; 2010 [cited 2019 Sep 8];87:41–52. Available from: http://link.springer.com/10.1007/s00253-010-2530-6
15. Barreiro C, Martín JF, García-Estrada C. Proteomics shows new faces for the old penicillin producer Penicillium chrysogenum. J Biomed Biotechnol [Internet]. 2012 [cited 2019 Sep 1];2012:1–15. Available from: http://www.hindawi.com/journals/bmri/2012/105109/
16. Nielsen J, Keasling JD. Engineering Cellular Metabolism. Cell [Internet]. Elsevier Ltd; 2016;164:1185–97. Available from: http://dx.doi.org/10.1016/j.cell.2016.02.004
17. Kamm B, Kamm M. Principles of biorefineries. Appl Microbiol Biotechnol [Internet]. 2004 [cited 2019 Sep 1];64:137–45. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14749903
18. Roddy DJ. Biomass in a petrochemical world. Interface Focus [Internet]. The Royal Society; 2013 [cited 2019 Sep 1];3:20120038. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24427511
19. Kamm B, Gruber PR, Kamm M, Pye EK. Biorefineries-Industrial Processes and Products: Status Quo and Future Directions [Internet]. Biorefineries-Industrial Process. Prod. Status Quo Futur. Dir. Wiley-VCH; 2008 [cited 2019 Sep 1]. Available from: https://www.wiley.com/en-se/Biorefineries+Industrial+Processes+and+Products:+Status+Quo+and+Future+Directions-p-9783527329533
20. Sandun Fernando *, Sushil Adhikari, Chauda Chandrapal and, Murali N. Biorefineries: Current Status, Challenges, and Future Direction. American Chemical Society ; 2006 [cited 2019 Sep 1]; Available from: https://pubs.acs.org/doi/abs/10.1021/ef060097w
21. Fengel D, Wegener G. Wood : chemistry, ultrastructure, reactions [Internet]. Walter de Gruyter; 1989 [cited 2019 Sep 2]. Available from:

https://books.google.se/books/about/Wood.html?id=x1B4uITKnt0C&redir_esc=y
22. Jönsson LJ, Martín C. Pretreatment of lignocellulose: Formation of inhibitory by-products and strategies for minimizing their effects [Internet]. Bioresour. Technol. Elsevier Ltd; Jan 1, 2016 p. 103–12. Available from: https://www.sciencedirect.com/science/article/pii/S0960852415014042#b0100
23. Biomass as a Resource for the Chemical Industry. https://www.ifeu.de/landwirtschaft/pdf/VCI_IFEU_Biomass_Chemical_Industry.pdf [Internet]. 2007 [cited 2019 Aug 30]; Available from: https://www.ifeu.de/landwirtschaft/pdf/VCI_IFEU_Biomass_Chemical_Industry.pdf
24. Sun N, Liu H, Sathitsuksanoh N, Stavila V, Sawant M, Bonito A, et al. Production and extraction of sugars from switchgrass hydrolyzed in ionic liquids. Biotechnol Biofuels [Internet]. BioMed Central; 2013 [cited 2017 Oct 13];6:39. Available from: http://biotechnologyforbiofuels.biomedcentral.com/articles/10.1186/1754-6834-6-39
25. Ruegg TL, Kim E-M, Simmons B a, Keasling JD, Singer SW, Lee TS, et al. An auto-inducible mechanism for ionic liquid resistance in microbial biofuel production. Nat Commun [Internet]. Nature Publishing Group; 2014;5:3490. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24667370
26. Stephanopoulos G. Synthetic Biology and Metabolic Engineering. ACS Synth Biol [Internet]. 2012 [cited 2019 Aug 29];1:514–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23656228
27. Bailey JE. Toward a science of metabolic engineering. Science [Internet]. American Association for the Advancement of Science; 1991 [cited 2019 Sep 2];252:1668–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2047876
28. Sauer U. Evolutionary Engineering of Industrially Important Microbial Phenotypes. Springer, Berlin, Heidelberg; 2001 [cited 2019 Sep 10]. p. 129–69. Available from: http://link.springer.com/10.1007/3-540-45300-8_7
29. Phaneuf PV, Gosting D, Palsson BO, Feist AM. ALEdb 1.0: A Database of Mutations from Adaptive Laboratory Evolution. bioRxiv [Internet]. Cold Spring Harbor Laboratory; 2018 [cited 2019 Jun 4];320747. Available from: https://www.biorxiv.org/content/10.1101/320747v1.full
30. Sandberg TE, Salazar MJ, Weng LL, Palsson BO, Feist AM. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. Metab Eng [Internet]. Elsevier Inc.; 2019;56:1–16. Available from:

https://linkinghub.elsevier.com/retrieve/pii/S1096717619301533
31. Portnoy VA, Bezdan D, Zengler K. Adaptive laboratory evolution-harnessing the power of biology for metabolic engineering. Curr Opin Biotechnol [Internet]. Elsevier Ltd; 2011;22:590–4. Available from: http://dx.doi.org/10.1016/j.copbio.2011.03.007
32. R. Lennen, K. Jensen, S. Malla, E. Mohamed, K. Chekina, M. Sommer, A. Feist AN and MH. Adaptive laboratory evolution of product-tolerant hosts for biobased chemical production [Internet]. 2016 SIMB Annual Meeting and Exhibition; 2016 [cited 2016 Aug 29]. Available from: https://sim.confex.com/sim/2016/webprogram/Paper32754.html
33. Luo H, Hansen ASL, Yang L, Schneider K, Kristensen M, Christensen U, et al. Coupling S-adenosylmethionine–dependent methylation to growth: Design and uses. Khosla C, editor. PLOS Biol [Internet]. Public Library of Science; 2019 [cited 2019 Sep 10];17:e2007050. Available from: http://dx.plos.org/10.1371/journal.pbio.2007050
34. Petzold CJ, Chan LJG, Nhan M, Adams PD. Analytics for Metabolic Engineering. Front Bioeng Biotechnol [Internet]. Frontiers; 2015 [cited 2019 Sep 10];3:135. Available from: http://journal.frontiersin.org/Article/10.3389/fbioe.2015.00135/abstract
35. LaCroix RA, Sandberg TE, O'Brien EJ, Utrilla J, Ebrahim A, Guzman GI, et al. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal medium. Appl Environ Microbiol [Internet]. American Society for Microbiology; 2015 [cited 2016 Aug 30];81:17–30. Available from: http://aem.asm.org/cgi/doi/10.1128/AEM.02246-14
36. Sandberg TE, Lloyd CJ, Palsson BO, Feist AM. Laboratory Evolution to Alternating Substrate Environments Yields Distinct Phenotypic and Genetic Adaptive Strategies. Kivisaar M, editor. Appl Environ Microbiol [Internet]. 2017 [cited 2017 Nov 22];83:AEM.00410-17. Available from: http://aem.asm.org/lookup/doi/10.1128/AEM.00410-17
37. Cardoso J. Development and application of computer-aided design methods for cell factory optimization [Internet]. Technical University of Denmark (DTU); 2017 [cited 2019 Sep 10]. Available from: https://orbit.dtu.dk/ws/files/140300426/thesis.pdf
38. Davy AM, Kildegaard HF, Andersen MR. Cell Factory Engineering. Cell Syst [Internet]. 2017 [cited 2019 Sep 16];4:262–75. Available from: https://linkinghub.elsevier.com/retrieve/pii/S2405471217300820
39. Burdock GA, Carabin IG. Generally recognized as safe (GRAS): history and description. Toxicol Lett [Internet]. 2004 [cited 2019 Sep 16];150:3–18. Available from:

https://linkinghub.elsevier.com/retrieve/pii/S0378427404000347
40. Yang ST. Bioprocessing-from Biotechnology to Biorefinery. Bioprocess Value-Added Prod from Renew Resour [Internet]. Elsevier; 2007 [cited 2019 Sep 9]. p. 1–24. Available from: https://www.sciencedirect.com/science/article/pii/B9780444521149500025
41. Rando OJ, Verstrepen KJ. Timescales of genetic and epigenetic inheritance. Cell [Internet]. 2007 [cited 2019 Sep 4];128:655–68. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0092867407001213
42. Jeong H, Lee SJ, Kim P. Procedure for Adaptive Laboratory Evolution of Microorganisms Using a Chemostat. J Vis Exp [Internet]. 2016 [cited 2019 Sep 4];e54446. Available from: http://www.jove.com/video/54446/procedure-for-adaptive-laboratory-evolution-microorganisms-using
43. Dragosits M, Mattanovich D. Adaptive laboratory evolution – principles and applications for biotechnology. Microb Cell Fact [Internet]. BioMed Central; 2013 [cited 2019 Sep 9];12:64. Available from: http://microbialcellfactories.biomedcentral.com/articles/10.1186/1475-2859-12-64
44. Alonso A, Campanario E, Martínez JL. Emergence of multidrug-resistant mutants is increased under antibiotic selective pressure in Pseudomonas aeruginosa. Microbiology [Internet]. 1999 [cited 2019 Sep 9];145:2857–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10537207
45. Lenski RE, Rose MR, Simpson SC, Tadler SC. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. Am Nat [Internet]. University of Chicago Press; 2002 [cited 2019 May 20];138:1315–41. Available from: https://www.journals.uchicago.edu/doi/10.1086/285289
46. Weikert C, Sauer U, Bailey JE. Use of a glycerol-limited, long-term chemostat for isolation of Escherichia coli mutants with improved physiological properties. Microbiology. Microbiology Society; 1997;143:1567–74.
47. Hong KK, Vongsangnak W, Vemuri GN, Nielsen J. Unravelling evolutionary strategies of yeast for improving galactose utilization through integrated systems level analysis. Proc Natl Acad Sci U S A. 2011;108:12179–84.
48. Carroll SM, Marx CJ. Evolution after Introduction of a Novel Metabolic Pathway Consistently Leads to Restoration of Wild-Type Physiology. Schneider D, editor. PLoS Genet [Internet]. 2013 [cited 2019 Dec 19];9:e1003427. Available from: https://dx.plos.org/10.1371/journal.pgen.1003427
49. Mohamed ET, Mundhada H, Landberg J, Cann I, Mackie RI, Nielsen AT, et al. Generation of an E. coli platform strain for improved sucrose utilization using adaptive laboratory evolution. Microb

Cell Fact [Internet]. BioMed Central; 2019;18:116. Available from: https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-019-1165-2
50. Zorraquino-Salvo V, Quinones-Soto S, Kim M, Rai N, Tagkopoulos I. Deciphering the genetic and transcriptional basis of cross-stress responses in Escherichia coli under complex evolutionary scenarios. bioRxiv [Internet]. Cold Spring Harbor Laboratory; 2014 [cited 2019 Dec 19];010595. Available from: https://www.biorxiv.org/content/early/2014/10/24/010595
51. Sandberg TE, Pedersen M, LaCroix RA, Ebrahim A, Bonde M, Herrgard MJ, et al. Evolution of Escherichia coli to 42°C and Subsequent Genetic Engineering Reveals Adaptive Mechanisms and Novel Mutations. Mol Biol Evol [Internet]. Oxford University Press; 2014 [cited 2019 Jun 4];31:2647–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25015645
52. González-Ramos D, Gorter De Vries AR, Grijseels SS, Van Berkum MC, Swinnen S, Van Den Broek M, et al. A new laboratory evolution approach to select for constitutive acetic acid tolerance in Saccharomyces cerevisiae and identification of causal mutations. Biotechnol Biofuels. BioMed Central Ltd.; 2016;9.
53. Bachmann H, Starrenburg MJC, Molenaar D, Kleerebezem M, Van Hylckama Vlieg JET. Microbial domestication signatures of Lactococcus lactis can be reproduced by experimental evolution. Genome Res. 2012;22:115–24.
54. Henson WR, Campbell T, DeLorenzo DM, Gao Y, Berla B, Kim SJ, et al. Multi-omic elucidation of aromatic catabolism in adaptively evolved Rhodococcus opacus. Metab Eng. Academic Press Inc.; 2018;49:69–83.
55. Nielsen J. Yeast cell factories on the horizon. Science (80-) [Internet]. 2015 [cited 2019 Sep 2];349:1050–1. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.aad2081
56. Li C, Tanjore D, He W, Wong J, Gardner JL, Sale KL, et al. Scale-up and evaluation of high solid ionic liquid pretreatment and enzymatic hydrolysis of switchgrass. Biotechnol Biofuels [Internet]. 2013;6:154. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3817576&tool=pmcentrez&rendertype=abstract
57. Li C, Knierim B, Manisseri C, Arora R, Scheller H V., Auer M, et al. Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification. Bioresour Technol [Internet]. Elsevier Ltd; 2010;101:4900–6. Available from: http://dx.doi.org/10.1016/j.biortech.2009.10.066
58. Bevan MW, Franssen MCR. Investing in green and white biotech. Nat Biotechnol [Internet].

Nature Publishing Group; 2006 [cited 2019 Jun 4];24:765. Available from: http://dx.doi.org/10.1038/nbt0706-765
59. Koutinas AA, Wang R, Webb C. Evaluation of wheat as generic feedstock for chemical production. Ind Crops Prod [Internet]. Elsevier; 2004 [cited 2018 Oct 4];20:75–88. Available from: https://www.sciencedirect.com/science/article/pii/S092666900400010X
60. Phaneuf P V., Gosting D, Palsson BO, Feist AM. Aledb 1.0: A database of mutations from adaptive laboratory evolution experimentation. Nucleic Acids Res. Oxford University Press; 2019;47:D1164–71.
61. Hill J, Nelson E, Tilman D, Polasky S, Tiffany D. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proc Natl Acad Sci U S A [Internet]. 2006 [cited 2019 Sep 11];103:11206–10. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0604600103
62. U.S. Department of Energy. U.S. Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry | Department of Energy [Internet]. Oak Ridge Natl. Lab. Oak Ridge, TN (United States); 2011 Aug. Available from: http://www.osti.gov/servlets/purl/1023318-ETXjs9/
63. Wang S, Cheng G, Dong J, Tian T, Lee TS, Mukhopadhyay A, et al. Tolerance Characterization and Isoprenol Production of Adapted Escherichia coli in the Presence of Ionic Liquids. ACS Sustain Chem Eng. 2019;7:1457–63.
64. Mundhada H, Schneider K, Christensen HB, Nielsen AT. Engineering of high yield production of L-serine in Escherichia coli. Biotechnol Bioeng. 2016;113:807–16.
65. Utrilla J, O'Brien EJ, Chen K, McCloskey D, Cheung J, Wang H, et al. Global Rebalancing of Cellular Resources by Pleiotropic Point Mutations Illustrates a Multi-scale Mechanism of Adaptive Evolution. Cell Syst [Internet]. Elsevier Inc.; 2016 [cited 2019 Jun 4];2:260–71. Available from: http://dx.doi.org/10.1016/j.cels.2016.04.003

List of publications

Included in this thesis

1. Generation of an E. coli platform strain for improved sucrose utilization using adaptive laboratory evolution. **Mohamed ET**, Mundhada H, Landberg J, Cann I, Mackie RI, Nielsen AT, et al. Microb Cell Fact. BioMed Central; 2019; 18:116.

2. Generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution. **Mohamed ET**, Wang S, Lennen RM, Herrgård MJ, Simmons BA, Singer SW, et al. Microb Cell Fact. BioMed Central; 2017; 16:1–15.

3. Generation of platform strains for aromatic acids utilizations using adaptive laboratory evolution. **Elsayed T. Mohamed**, Davinia Salvachúa, Christine Singer, Mohammad S. Radi, Markus J. Herrgård, Gregg T. Beckham, Adam M. Feist (2019), Manuscript in preparation.

List of abbreviations

ALE	Adaptive laboratory evolution
Whole genome re-sequencing	WGS
SUC	Sucrose
FRU	Fructose
GLUC	Glucose
Ionic liquids	ILs
TALE	Tolerance adaptive laboratory evolution
[C2C1Im][OAc]	1-ethyl-3-methylimidazolium acetate
[C4C1Im]Cl	1-butyl-3-methylimidazolium chloride
p-CA	*p*-coumaric acid
FA	Ferulic acid
NGS	Next generation sequencing
CSC	Sucrose integrated cassette

The following section presents the work done and detailed discussions of the generated results in this thesis presented in three papers, two of which are published and the third is under preparation.

Paper 1: Generation of an *E. coli* platform strain for improved sucrose utilization using adaptive laboratory evolution

Microbial Cell Factories

RESEARCH Open Access

Generation of an *E. coli* platform strain for improved sucrose utilization using adaptive laboratory evolution

Elsayed T. Mohamed[1], Hemanshu Mundhada[1], Jenny Landberg[1], Isaac Cann[2], Roderick I. Mackie[2], Alex Toftgaard Nielsen[1], Markus J. Herrgård[1] and Adam M. Feist[1,3*]

Abstract

Background: Sucrose is an attractive industrial carbon source due to its abundance and the fact that it can be cheaply generated from sources such as sugarcane. However, only a few characterized *Escherichia coli* strains are able to metabolize sucrose, and those that can are typically slow growing or pathogenic strains.

Methods: To generate a platform strain capable of efficiently utilizing sucrose with a high growth rate, adaptive laboratory evolution (ALE) was utilized to evolve engineered *E. coli* K-12 MG1655 strains containing the sucrose utilizing *csc* genes (*cscB*, *cscK*, *cscA*) alongside the native sucrose consuming *E. coli* W.

Results: Evolved K-12 clones displayed an increase in growth and sucrose uptake rates of 1.72- and 1.40-fold on sugarcane juice as compared to the original engineered strains, respectively, while *E. coli* W clones showed a 1.4-fold increase in sucrose uptake rate without a significant increase in growth rate. Whole genome sequencing of evolved clones and populations revealed that two genetic regions were frequently mutated in the K-12 strains; the global transcription regulatory genes *rpoB* and *rpoC,* and the metabolic region related to a pyrimidine biosynthetic deficiency in K-12 attributed to *pyrE* expression. These two mutated regions have been characterized to confer a similar benefit when glucose is the main carbon source, and reverse engineering revealed the same causal advantages on M9 sucrose. Additionally, the most prevalent mutation found in the evolved *E. coli* W lineages was the inactivation of the *cscR* gene, the transcriptional repression of sucrose uptake genes.

Conclusion: The generated K-12 and W platform strains, and the specific sets of mutations that enable their phenotypes, are available as valuable tools for sucrose-based industrial bioproduction in the facile *E. coli* chassis.

Keywords: *Escherichia coli*, Renewable feedstocks, Sucrose, Adaptive laboratory evolution, Platform strains

Introduction

There is a significant interest in the utilization of renewable carbon feedstocks for bioprocesses due to both environmental and economic factors. Sucrose from sugarcane can serve as a renewable carbon source as it originates from a cheap abundant feedstock source which can lower carbon source costs [1, 2], and it can be readily utilized by biological systems in fermentation processes either from the raw source as sugarcane juice [3] or from the refined byproduct of the sugar industry as molasses [4, 5].

The ability of *Escherichia coli*, an industrial biotechnology workhorse [6–8], to utilize sucrose as a sole carbon source depends on the specific *E. coli* strain used [9]. *E. coli* W is the only well-known *E. coli* strain generally regarded as safe that can utilize sucrose as a carbon source and can grow robustly on it when compared to other carbon sources, such as glucose [10]. The genetic basis and molecular control of sucrose metabolism in *E. coli* W (as well as *E. coli* EC3132) have been characterized [11, 12] and this knowledge is the basis

*Correspondence: afeist@ucsd.edu
[3] Department of Bioengineering, University of California, 9500 Gilman Drive La Jolla, San Diego, CA 92093, USA
Full list of author information is available at the end of the article

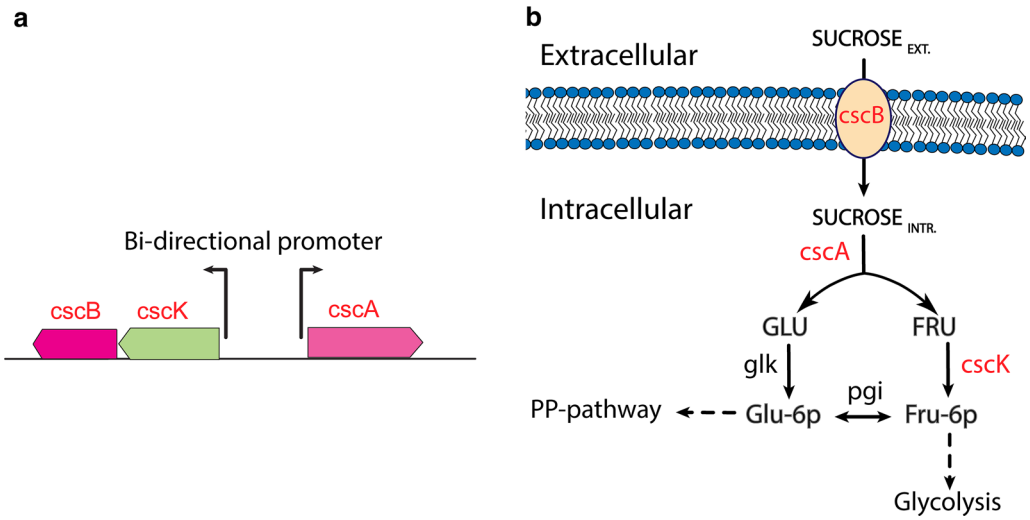


Fig. 1 Schematic presentation of sucrose utilization mechanism with gene cassette integrated into *E. coli* K-12 MG1655 and route of sucrose entering into the central carbon metabolism. **a** The sucrose integrated cassette *csc* consists of *cscA*, *cscK*, and *cscB*. CscBK and *cscA* expression is derived by a bi-directional promoter. **b** Sucrose utilization starts with uptake of sucrose across the cell membrane by *cscB* (sucrose permease). Inside the cell sucrose is split into glucose and fructose by *cscA* (invertase). Glucose and fructose are then phosphorylated by glucokinase (glk) and fructokinase (*cscK*) respectively. Glucose-6-phosphate can be converted to fructose-6-phosophate by the isomerase (pgi) or enter directly into the pentose phosphate pathway (PP-pathway) or Entner-Doudoroff pathway (ED-pathway); fructose-6-phosphate enters the glycolysis directly

for constructing sucrose utilizing *E. coli* strains for the current study. Engineering non-native sucrose utilization into additional *E. coli* strains has been investigated, but in all cases, the generated strains were characterized by slow growth rates on sucrose as compared to glucose. Such low growth rates would limit productivity in industrial processes, a key component for a viable bioprocess [13, 14]. Most of the previous approaches to engineer *E. coli* to use sucrose as a carbon source were hampered by slow growth rates and phenotypic instability due to unstable plasmid systems that may require antibiotic addition for stability [15] or due to a high burden exerted on the cell by high copy number plasmids [11, 16–18]. For example, the K-12 strains generated by Tsunekawa et al. using chromosomal integration grew with very slow rates on sucrose and random mutagenesis was required to improve the growth rate modestly [19]. Another approach examined was to chromosomally integrate sucrose utilization genes from *E. coli* W into an *E. coli* K-12 strain [20]. The K-12 strain generated was able to grow on sucrose, but the growth rate was 30% lower than that on glucose. Thus, there is still a need to efficiently engineer sucrose metabolism in multiple *E. coli* strains, each of which have their own strain-specific advantages for host selection for viable bioprocesses [6]. Such host selection factors to consider include product tolerance, phage resistance, the native metabolic flux distribution either towards a native or heterologous production pathway, transformation efficiency, convenience to perform metabolic changes, and ease to scale-up [21].

In order to demonstrate the ability to generate additional *E. coli* strains that can efficiently consume sucrose, genetic constructs were designed based on the chromosomally encoded sucrose catabolism operon, *csc*, from *E. coli* W and then subjected to adaptive laboratory evolution (ALE). The *csc* gene cluster containing the *cscB* gene that encodes sucrose permease, *cscK* that encodes a fructokinases, and *cscA* that encodes sucrose-6-phosphate hydrolase (invertase), was integrated into the *E. coli* K-12 MG1655 chromosome [11] (Fig. 1). The *cscR* gene that encodes for a *csc*-specific repressor and negatively controls the expression of the *csc* regulon was not integrated to allow constitutive expression. The engineered strain was then optimized using ALE to generate multiple evolved *E. coli* strains able to metabolize sucrose with fast growth comparable to that on glucose. Additionally, the approach to analyze multiple independent populations and multiple independent isolates from each population made it possible to effectively reveal key causal mutations by comparing the independent lineages and focusing on instances of parallel evolution. This approach also likely identified a broader landscape of mutations (e.g., multiple alleles of the same gene) as compared to sequencing a single lineage and reverse engineering all of the mutations to find causality. The ALE approach used in the current work used an automated platform with several parallel replicates as previously described [22, 23].

Overall, cells were repeatedly grown in a batch aerobic cultivation mode with passage in the exponential phase, which applies selection pressure for rapid growth rate. At the end of the evolution, populations and clones from the ALE derived endpoints were characterized in terms of their genome sequence, growth rate, carbon uptake rate, and growth yield. Furthermore, the evolved strains were characterized on sugarcane juice as a substrate demonstrating their ability to use this renewable feedstock efficiently.

Methods and materials

Media and reagents

M9 minimal medium

M9 sucrose medium contained 20 g/L sucrose unless stated otherwise, 1× M9 salts, 2 mM $MgSO_4$, 100 µM $CaCl_2$ and 1× trace elements and Wolfe's vitamin solution. M9 salts, trace elements and vitamins were prepared in concentrated stocks. Stock M9 salts solution consisted of 10 × 68 g/L Na_2HPO_4 anhydrous, 30 g/L KH_2PO_4, 5 g/L NaCl, and 10 g/L NH_4Cl dissolved in Milli-Q filtered water. M9 trace elements stock was a 2000x solution with composition of 3.0 g/L $FeSO_4{\cdot}7H_2O$, 4.5 g/L $ZnSO_4{\cdot}7H_2O$, 0.3 g/L $CoCl_2{\cdot}6H_2O$, 0.4 g/L $Na_2MoO_4{\cdot}2H_2O$, 4.5 g/L $CaCl_2{\cdot}H_2O$, 0.2 g/L $CuSO_4{\cdot}2H_2O$, 1.0 g/L H_3BO_3, 15 g/L disodium ethylenediamine-tetra-acetate, 0.1 g/L KI, 0.7 g/L $MnCl_2{\cdot}4H_2O$ and concentrated HCL dissolved in Milli-Q filtered water. Vitamin mix and trace elements concentration was [1×] in the M9 medium.

Sugarcane juice medium

Raw sugarcane juice was collected from a sugar cane mill (Enterprise Factory, Patoutville, Louisiana, USA) and transported to the University of Illinois Energy Farm where it was stored at − 80 °C until required for further use. Sugarcane juice (SCJ) for the cultivation medium was prepared by centrifugation in order to remove solid particles (rpm ≥ 10,000×*g* for 20 min at 4 °C) then filter sterilized with 0.2 µm pore size filter. Later, Sugarcane juice (SCJ) medium was prepared by mixing clarified sugarcane juice with M9 salts (final concentration 1×), 2 mM $MgSO_4$, 100 µM $CaCl_2$ and 1× trace elements and Wolfe's vitamin solution. HPLC analysis of the clarified sugar cane juice showed the following sugar composition (g/L): sucrose 231.8, glucose 3.5; fructose 3.0.

Strains constructions with cscBKA cassette using homologous recombination

General recombinant DNA techniques were applied according to standard protocols for one-step cloning and chromosomal integration of DNA (pOSIP), Additional file 2: Text file S1 [24, 25]. PCR products were purified using plasmid miniprep kit, QIAprep Spin Miniprep Kit. Here, the integration site of *csc* DNA sequence was into *E. coli* 186 attB site using pOSIP-KO integration module. The integration site of DNA sequence into bacterial chromosomes was selected based on the available integration sites. Csc gene cluster (*cscK*, *cscB*, *cscA*) (Fig. 1a) was amplified from donor strain, *E. coli* W using primers; Forward primer 5′-ATGCATCUGGGATATAGAGCT ATCGACAACAACCG-3′ and Reverse primer 5′-AGA GGGCUTTATGTTAACCCAGTAGCCAGAGTGCTC -3′. *E. coli* K-12 construct with *csc* cassette integrated was abbreviated to MGcscBKA. MGcscBKA derivative with a variant SNP in *cscB* Q353H was abbreviated MGcscB-KAp [16].

Adaptive laboratory evolution of developing improved fitness in start strains

Three biological replicates from each of *E. coli* strains; two genetic constructs of K-12 MG1655 with csc gene cassette MGcscBKA and MGcscBKAp and *E. coli* W strain, were grown overnight in M9 minimal media with 20 g/L sucrose. On the next day 150 µL of the overnight culture was passed into a new fresh 30 mL tube filled with a total working volume of 15 mL M9 medium (i.e., a 1:100 ratio) with 20 g/L sucrose as the sole carbon source. Bacterial cells were serially passaged during exponential growth phase for approximately 40 days using an automated liquid-handler platform as described by Lacroix et al. [22]. The liquid-handler platform was utilized to automate executing multiple evolution experiments at the same time. It has the capacity to aspirate cell cultures when passing from a mature culture tube to a freshly filled tube based on a pre-defined media recipe in a sterile strictly-controlled platform, in an effort to keep the cells growing and passing at the exponential phase of growth. The commonly encountered exponential phase was from the time of the inoculation to an approximate optical density of 600 nm (OD_{600nm}) of 1.3–2.0 in order to keep the cells under a constant selection pressure. Cells were cultured in a heat block at 37 °C with magnetic stir bar for full aeration at 1200 rpm. Periodically, OD_{600nm} was measured at a time determined by a predictive script and once OD_{600nm} reached approximately 1.3 using Sunrise plate reader (Tecan, Männedorf, Switzerland), 150 µL was passed into a tube with a 15 mL working volume of fresh media. The common conversion factor between the plate reader used and the Benchtop spectrophotometer is 4.2. This process was repeated until a significant increase in fitness was achieved. Periodically, glycerol cryogenic stocks were prepared and stored at − 80 °C for any culture restarting. Endpoint population samples were streaked on agar plates to select single clones for whole genome resequencing.

Genome sequencing and mutation calling of the ALE derived strains

Overall, there were six endpoint populations for each of K-12 MG1655 constructs and 3 endpoints populations from *E. coli* W were selected along with clones derived from for whole genome re-sequenced in order to reveal their underlying genotypes. Genomic DNA was extracted from overnight cultures at the stationary phase of growth using PureLink® Genomic DNA extraction kits (Invitrogen, CA). The quality of extracted DNA was assessed by using a Nanodrop spectrophotometer. Concentration of the extracted DNA was quantified using Qubit ds-DNA high sensitivity assay. Paired-end re-sequencing libraries were generated using a 300 cycle (150 bp × 2) kit from Illumina (San Diego, CA) with loading concentration of 1.2 pico-Molar on Illumina Nextseq sequencer (Model 550). Mutation finding was performed using a pipeline as described in Phaneuf et al. [26] based on *Breseq* version 0.30.1 [27] to map sequenced reads to the reference strain (NCBI accession number NC_000913.3, K-12 MG1655 and NC_017664, W). The average coverage for each of the resequenced samples was over 25x. For population samples sequenced, mutations were reported if they were over 20% frequency unless they were found in a clone isolated from a given population sample. In this case, the frequency of a clone mutation was reported, if found, in the population (see Additional file 1: Data file S1).

Validation of mutations causality in *rpoB, rpoC* and *pyrE-rph* genes

Identified causal mutations found in the current study (see mutation analysis results) and the control reference ALE experiment by LaCroix et al. [22] on glucose (GLU-ALE) were used to check causality in the identified genes or genetic regions on either sucrose and glucose minimal media. Briefly, the identified key single point mutations (SNPs) in open reading frames (ORF) mainly for *rpoB*, *rpoC* genes and the unique deletion in the intergenic region between *pyrE*/*rph* were used to check the causality of mutations in these specified genetic regions with sucrose for growth increase from the two experiments. Accordingly, selected isolates from ALE derived clones on sucrose which harbor any single mutation or double mutations in these genes were selected as well. Fitnesses increase comparison relative to the starting strain were examined to investigate the effect/essentiality of the key causal mutations for fast growth on either sucrose or glucose as a sole carbon source. The same recombinant DNA technique used to generate MGcscBKA constructs with the *csc* regulon was applied here to generate GLU-ALE constructs with a sucrose utilization cassette (GLU-ALE_csc constructs).

Extracellular metabolites and physiological properties

Cultures of the re-sequenced clones were inoculated from stationary phase overnight cultures into media M9 containing sucrose under the same conditions as the ALE experiment. Samples were aliquoted over the growth curve to measure optical density OD_{600nm} and collect extracellular metabolites. Extracellular metabolites were collected as supernatant from each growing culture using 0.2 µm filter to remove the cells. Supernatants were collected and saved at −20 °C for subsequent chromatographic analyses. Concentration of sugars (glucose, fructose and sucrose) beside other organic acids were analyzed using high performance liquid chromatography (HPLC) column (UltiMate 3000, Thermo-Fischer Scientific, Waltham, Massachusetts, USA). The metabolites were separated using an Aminex HPX-87H ion exclusion column (Bio-Rad, Hercules, California, USA) and were isocratically eluted at 30 °C, with a flow rate of 0.6 mL/min, using a 5 mM sulfuric acid solution as mobile phase. The refractive index (RI) detector was selected for detection. Sample concentrations were quantified by comparing to a standard curve of known concentrations. Substrate uptake and metabolites excretion rates were calculated from multiplying the growth rate and the slope of a linear regression of gram dry cell weigh (gCDW) versus the substrate or products concentration. The biomass yield at the steady state (Y_{X/S_ss}) was calculated as the quotient of the growth rate and the sucrose uptake rates during the exponential growth phase.

Results

Evolution of multiple *E. coli* strains to grow rapidly on sucrose minimal media

ALE was utilized to generate strains with improved fitness utilizing sucrose as a sole carbon source. Three different *E. coli* starting strains were used; two engineered K-12 MG1655 strains with sucrose utilization enabling csc constructs and wild-type *E. coli* W. The two constructs inserted into the K-12 MG1655 host genome differed by a SNP mutation in the *cscB* gene, resulting in the residue change Q353H, and were labeled as MGcscBKA and MGcscBKAp (see Additional file 2: Text S1 for a detailed description). The Q353H derivative of *cscB* has been demonstrated to display an increased sucrose uptake rate in *E. coli* EC3132 [16], thus it was reasoned that it may have an impact when heterologously expressed in K-12 MG1655.

Three independent populations of each starting strain were evolved in batch and in parallel under the strict selection pressure of continuous exponential growth for approximately 40 days (Table 1, Fig. 2a–c). Termination of the ALE experiment was determined based on two key parameters [28]; the delta change in growth rate (Δμ) and

Table 1 Properties of the ALE experiments end point populations

ALE experiment	Strain (population)	Total CCD x10^{12}	Ratio of fitness increase to start strain
K-12 MGcscBKA	1	8.65	1.72
	2	8.75	1.46
	3	8.94	1.61
K-12 MGcscBKAp	1	9.02	1.48
	2	9.02	1.47
	3	8.85	1.48
E. coli W	1	9.71	1.29
	2	9.78	1.31
	3	9.87	1.22

Ratios of fitness increase were determined from growth rates that were calculated based on the last three flasks during exponential growth. CCD, cumulative cell divisions

the passage size. Therefore, the experiments were terminated when the parallel replicates showed a small to no delta change in the growth rate similar to the control ALE on glucose [22], with the passage volume equal to 1%. Each of the replicate evolutions underwent between approximately 8.65×10^{12} to 9.87×10^{12} cumulative cell divisions (CCD, Table 1). The use of CCD has been demonstrated as a useful timescale for ALE experiments as a time coordinate as it accounts for variability due to a varying number of cells passed serially from one flask to the next [29].

The observed growth rate trajectories during the ALE experiments for each of the evolved populations are shown in Fig. 2 for the MGcscBAK, MGcscBKAp, and *E. coli* W strains. Each population displayed an increase in growth rate over the starting strain (Table 1). The growth rate increases were 1.49 ± 0.098-, $1.75 \pm < 0.01$-, 1.25 ± 0.036-fold faster than the starting strains for the MGcscBKA, MGcscBKAp and *E. coli* W strains, respectively (standard deviation, $n = 3$). There was one observed fitness jump (i.e., increase in growth rate) for all the populations across all independent ALE replicates. The two similar K-12 MGcscBKA and MGcscBKAp strains evolved to relatively similar growth rates, with similar dynamics along the course of the ALE, whereas the fold increase in fitness for *E. coli* W was small given its initial faster wild-type growth rate.

Physiological characterization of evolved clones

Clones from the endpoint populations of each of the nine replicate ALEs were isolated and characterized in terms of growth rate and compared to the starting strains to understand the enhanced evolved phenotype. One clone isolated from each replicate endpoint population was analyzed, three clones for each of the three replicates of

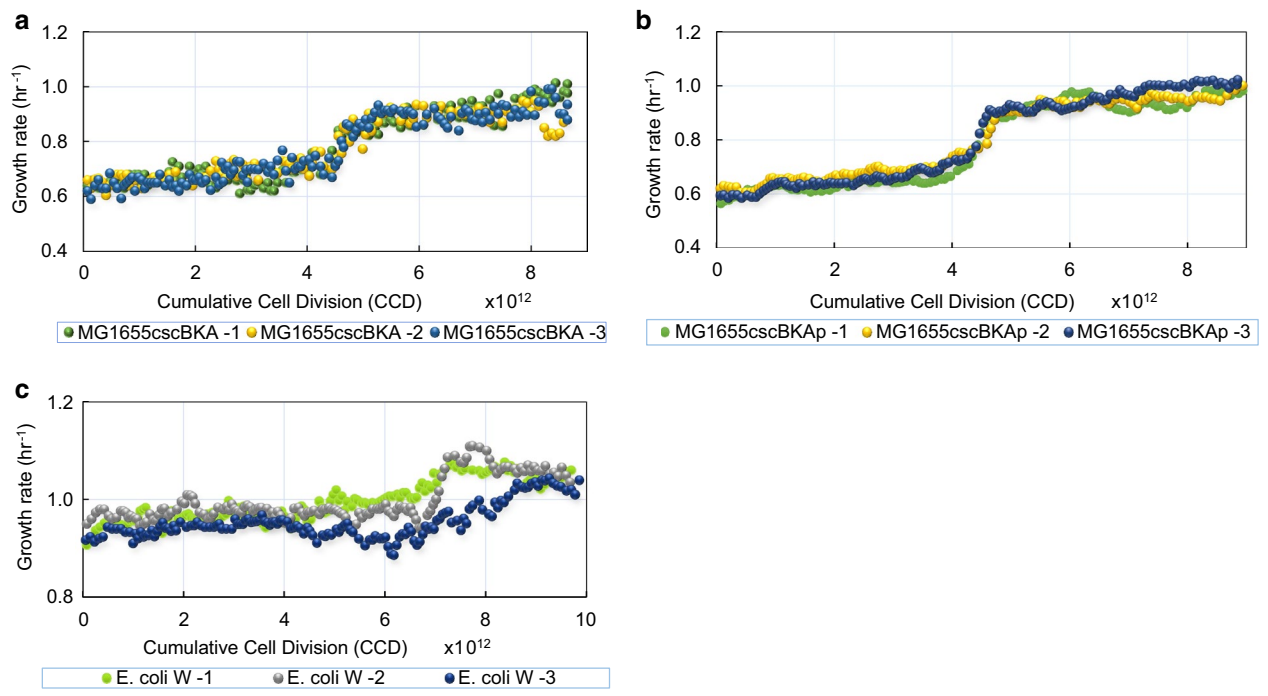


Fig. 2 Plots of population fitness (i.e., growth rate) trajectories during the ALE experiments on sucrose. Three biological replicates of each of the *E. coli* strains: **a** K-12 MG1655cscBKA, **b** K-12 MG1655cscBKAp, and **c** W were evolved on 20 g/L sucrose minimal medium. Shown are the growth rates versus cumulative cell divisions (CCD) for the three biological replicates for each strain

Table 2 Physiological data for the evolved isolates on sucrose M9 medium and sugarcane juice M9 minimal medium (SCJ medium)

Clone/ construct	Clone ID	M9 Sucrose medium	SCJ medium					
		Growth rate, μ (h^{-1}) on M9 2% sucrose	Growth rate, μ (h^{-1})	Final density (gCDW/L)	Sucrose uptake rate (mmol $gCDW^{-1}$ h^{-1})	Acetate production rate (mmol gDW^{-1} h^{-1})	Biomass yield, $Y_{X/S}$, (gCDW g of $sucrose^{-1}$)	Uptake rate fold increase vs start strain
K-12 MGcscBKA	Starting strain	0.54 ± 0.01	0.68 ± 0.03	1.47 ± 0.04	7.56 ± 0.96	4.20 ± 0.20	0.28 ± 0.01	1.00
	1	0.85 ± 0.01	0.88 ± 0.02	1.91 ± 0.03	8.71 ± 0.2	4.73 ± 0.11	0.30 ± 0.01	1.15
	2	0.63 ± 0.01	0.68 ± 0.02	1.52 ± 0.04	6.54 ± 0.23	5.67 ± 0.54	0.30 ± 0.01	0.86
	3	0.83 ± 0.01	0.86 ± 0.02	1.68 ± .0.03	7.84 ± 0.18	5.42 ± 0.13	0.32 ± 0.01	1.04
K-12 MGcscBKAp	Starting strain	0.64 ± 0.01	0.73 ± 0.01	1.48 ± 0.02	7.01 ± 1.9	4.99 ± 0.25	0.30 ± 0.01	
	1	0.86 ± 0.01	0.91 ± 0.01	1.38 ± 0.02	9.89 ± 0.19	6.83 ± 0.01	0.27 ± 0.01	1.41
	2	0.88 ± 0.01	0.88 ± 0.01	1.44 ± 0.02	9.50 ± 0.01	8.27 ± 0.01	0.27 ± 0.01	1.36
	3	0.80 ± 0.01	0.82 ± 0.01	1.51 ± 0.01	8.98 ± 0.11	7.30 ± 0.09	0.27 ± 0.01	1.28
E. coli W	Starting strain	0.90 ± 0.02	0.97 ± 001	1.81 ± 0.05	9.88 ± 0.35	6.60 ± 0.49	0.29 ± .01	
	1	1.10 ± 0.01	1.06 ± 0.05	1.68 ± 0.06	10.5 ± 0.47	15.90 ± 0.68	0.29 ± 0.012	1.06
	2	0.95 ± 0.01	0.96 ± 0.02	1.74 ± 0.05	14.10 ± 1.0	6.50 ± 0.51	0.20 ± 0.02	1.43
	3	0.95 ± 0.01	1.08 ± 0.02	1.92 ± 0.01	12.51 ± 1.1	7.40 ± 0.5	0.25 ± 0.03	1.27

The physiological properties of each of the clones isolated from the independent endpoint ALE experiments were compared to examine whether there were any improved phenotypic outcomes across the different experiments

MGcscBAK, MGcscBKA, and *E. coli* W. The increases in growth rates on sucrose M9 were between 1.17 and 1.57-fold for MGcscBKA clones, 1.25–1.38-fold for MGcscBKAp, and 1.06–1.22-fold for *E. coli* W, Table 2. Surprisingly, there was one isolated clone that had the same fitness, i.e. growth rate, as the starting strain from MGcscBKA (the '#2' clone) (Fig. 3). However, given that there were two additional clones from this starting strain from the additional two independent replicates with a significant improvement in fitness, additional clones were not examined.

Furthermore, growth profiling of all of the sucrose-evolved endpoint clones was performed in glucose M9 media. In agreement with the observed sucrose M9 phenotypes, all clones demonstrated an increased growth rate over the start strains without any significant changes to the final cellular density (gCDW/L) (Fig. 4). One exception to this was *E. coli* W replicate #2 which exhibited a slightly lower growth rate than the starting strain. The observation that the vast majority of the sucrose-evolved strains also displayed a fast growth phenotype on glucose is an indication that most mutations acquired during the sucrose ALE experiment were beneficial for growth on both sugars. Such a phenotype is advantageous for a platform strain as it could be used in multiple media conditions. All isolated clones were then sequenced to examine their genetic basis along with population sequencing to better understand the genotypes responsible for the observed increases in population fitness from all of the independent ALE experiments.

Whole genome sequencing and mutation analysis

Whole genome sequencing was used to determine the genetic basis of the improved fitness phenotypes for the evolved *E. coli* strains on sucrose minimal medium. The nine isolated endpoint clones were sequenced along with the corresponding populations from which they were isolated. Overall, there were nine genes or genetic regions identified from the K-12 strains found in clones and/or populations (frequency cutoff ≥ 0.20) that were mutated either once or several times across the parallel independent replicates. This number was five for *E. coli* W. Additionally, there was a range of 1-3 unique mutations found in all of the clones sequenced, with most clones possessing one mutation. When looking at the population data, there was a range of 0-3 mutations detected in each sample with a frequency cutoff ≥ 0.20, indicating different levels of clonal interference in different replicates. Key mutations were identified by comparing all clones and population samples sequenced and identifying open reading frames (i.e., genes), or intragenic regions that had multiple unique mutations or were mutated across independent experiments (Table 3). Interestingly, there were no shared key mutations found between the K-12 and *E. coli* W experiments. Full mutation lists for each sequenced sample are given in Additional file 1: Data file S1.

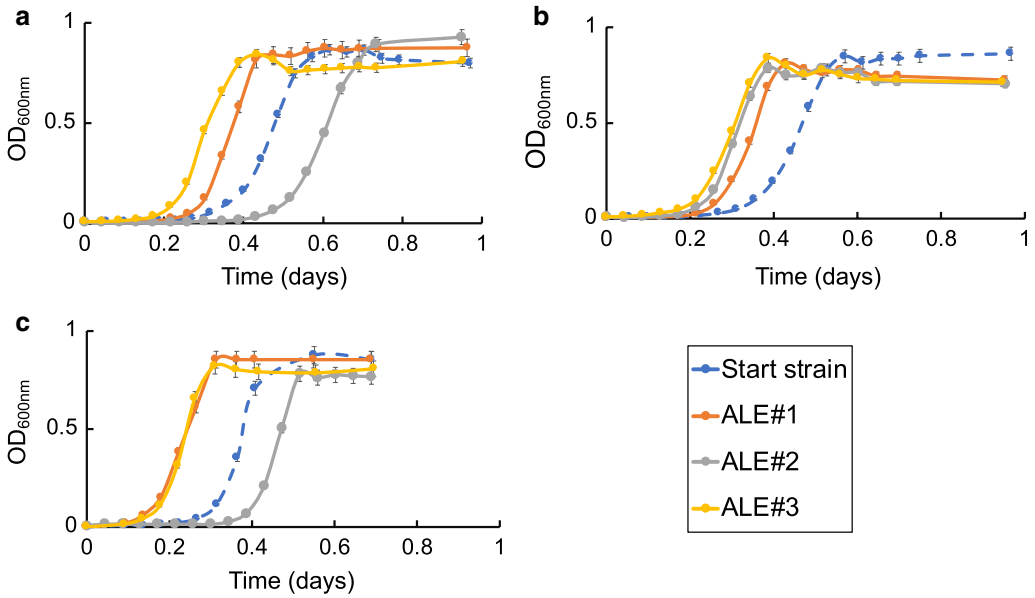


Fig. 3 Growth profiles of ALE clones isolated from the end populations of the evolution experiments versus starting strains on M9 minimal medium with 20 g/L sucrose. Error bars represent standard deviation (n = 3). MG1655 construct with *csc* (MGcscBKA) growth profile is represented on **a**, MGcscBKA derivatives with *cscB* SNP (MGcscBKAp) mutation on **b** and *E. coli* W strain on **c**. Optical density measurement at OD_{600} were measured on a plate reader and the common ratio between the plate reader OD_{600} and a benchtop spectrophotometer with a 1 cm path length is 4.2

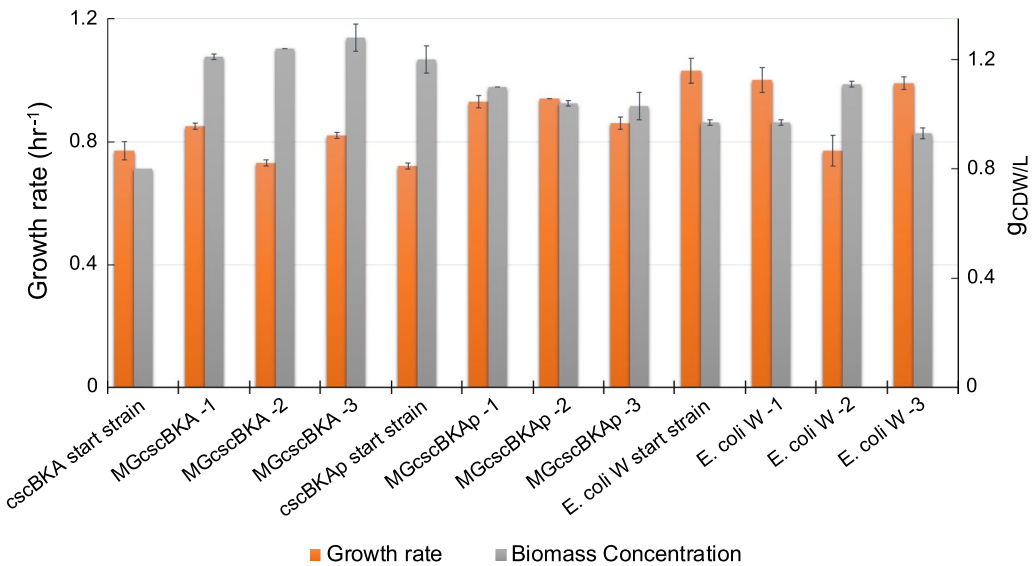


Fig. 4 Characterization of the isolated sucrose-evolved clones on 2 g/L glucose M9 minimal medium. Shown is a stacked histogram of growth rate (h^{-1}, orange) of the selected clones and their corresponding final cellular density (gCDW/L, grey). The plot shows that mutations acquired during the ALE experiment on sucrose were also largely beneficial, to a similar extent, on glucose as a sole carbon source. Error bars represent standard deviation of biological replicates (n = 3). Data is shown for both the starting strains and evolved clones for the three different sets of evolved strains

The most predominantly mutated genes or intergenic regions identified across multiple independent replicates in the evolved K-12 strains were related to the RNA polymerase subunits *rpoB* and *rpoC*, as well as orotate phosphoribosyltransferase, *pyrE*. The first key mutations were found in the β (beta) and β' (beta prime) subunit of RNA polymerase encoded in *rpoB* and *rpoC*, respectively [30, 31]. The number of unique mutations for each

Table 3 Key mutations found after evolution of *E. coli* on sucrose

Strain	Gene or genetic region	Mutation	Mutation type	Function	Number of independent occurrences	Sample ID (frequency of mutation in population samples)
K-12 starting strains MGcscBKA and MGcscBKAp	*pyrE/rph*	Δ82 bp intragenic	DEL	Orotate phosphoribosyltransferase/ribonuclease PH	2	MGcscBKA A1 F136 I0 (43%), MGcscBKA A1 F136 I1, MGcscBKA A3 F140 I0 (25%)
		T→G intergenic (−47/+19)	SNP		1	MGcscBKAp A3 F139 I0 (71%)
	rpoB	S621F (TCC→TTC)	SNP	RNA polymerase subunit β	1	MGcscBKA A1 F136 I0 (48%), MGcscBKA A1 F136 I1
		Q618L (CAG→CTG)	SNP		1	MGcscBKA A2 F134 I0 (4%)
	rpoC	T1045P (ACC→CCC)	SNP	RNA polymerase subunit β′	2	MGcscBKA A3 F140 I0 R1 (20%), MGcscBKA A3 F140 I1 R1, MGcscBKAp A3 F139 I0 (83%)
		R1075C (CGT→TGT)	SNP		2	MGcscBKAp A1 F138 I0 (93%), MGcscBKAp A1 F138 I1, MGcscBKAp A2 F140 I0 (100%), MGcscBKAp A2 F140 I1
		L770R (CTC→CGC)	SNP		1	MGcscBKA A3 F140 I0 (76%)
		R978C (CGT→TGT)	SNP		1	MGcscBKAp A3 F139 I0 (9.7%), MGcscBKAp A3 F139 I1
		Q665K (CAG→AAG)	SNP		1	MGcscBKA A2 F134 I0 (6%)
		R1174P (CGT→CCT)	SNP		1	MGcscBKA A2 F134 I0 (18%)
E. coli W	*cscR*	Δ1403 bp [cscR, dsdX, dsdA]	DEL	Sucrose operon repressor (Csc operon regulatory protein), permease DsdX, D-serine dehydratase (EC:4.3.1.18)	1	*E. coli* W A1 F155 I1
		Δ84 bp	DEL	Sucrose operon repressor (Csc operon regulatory protein)	1	*E. coli* W A3 F156 I1
		Δ10 bp	DEL		1	*E. coli* W A1 F155 I0 (74%)
		Δ21 bp	DEL	Sucrose operon repressor (Csc operon regulatory protein)	1	*E. coli* W A3 F156 I0 (46%)
	mrdB	S31R (AGC→AGG)	SNP	Rod shape-determining protein RodA	1	*E. coli* W A1 F155 I0 (77%)
		S37R (AGC→AGA)	SNP		1	*E. coli* W A2 F158 I0 (81)
		E270Q (GAA→CAA)	SNP		1	*E. coli* W A2 F158 I1
	mrdA	W434C (TGG→TGC)	SNP	Penicillin-binding protein 2	1	*E. coli* W A1 F155 I1

The sample ID has a unique identifier—(A) refers to the independent replicate, (F) refers to the flask number, and (I) to the type of the isolate: population (0) or clone (1)

of *rpoB* and *rpoC* found across parallel populations and clones was 2 and 6 unique SNPs, respectively. Both *rpoB* mutations occurred in one region between amino acids residues (AAR) 618 and 621, whereas the *rpoC* mutations occurred closer to the carboxyl terminus (AAR total 1342) of the beta prime subunit at AAR between 665 and 1174, Table 3. There was one co-occurrence of a *rpoC* mutation, T1045P (ACC→CCC), which was observed in experiments starting from both K-12 strains. Interestingly, mutations in *rpoB* and *rpoC* genes have previously been found repeatedly across all different K-12 ALE experiments, indicating a very high level of parallel evolution [22, 32, 33]. It is also interesting to note that no clone contained mutations in both genes, but there were often multiple mutations in these genes found in population samples (never adding up to more than approximately 1 in mutation frequency). Specific mutations in these two subunits of the RNA polymerase, i.e. the beta and the beta prime, were found to carry beneficial growth advantages when *E. coli* grows on minimal medium with a range of different carbon sources [22, 32–36]. The second key mutated region observed in K-12 strains were related to *pyrE* expression. K-12 strains are known to possess a frame shift in *rph* which leads to pyrimidine starvation on minimal media and can be alleviated by mutations [37]. Similar mutations were reported in different ALE studies on different carbon sources [32, 34]. Mutations in this genetic region were found to improve fitness on minimal media with an increase in growth rate of 17% over the starting strain on glucose [22].

Strain-specific adaptive mutations in *E. coli* W clones were predominantly affecting a metabolic regulation pathway targeting the *csc* operon and cell wall biosynthesis through rod shape determining proteins. Many of the mutations found related to the *csc* operon were in the *cscR* gene, which encodes a transcriptional repressor regulator for *csc* operon (*cscB*, *cscA*, *cscK* genes) in low concentrations of sucrose [11, 16]. A cscR mutation was observed in two endpoint clones (out of three) and a total of four mutations relating to this gene were found overall when considering population sequencing. The clonal cscR mutations were an intragenic in-frame Δ84 bp deletion and an Δ1403 bp deletion which also included the *dsdX* and *dsdA* genes located next to *cscR* on the chromosome. Both mutations are likely a disruption of the cscR gene with potentially a loss of function. The two genes *dsdX* and *dsdA* are pseudogenes coding for a D-serine transporter and D-serine ammonia-lyase, respectively, as part of serine degradation pathway [38, 39]. Deletions of *cscR* were previously demonstrated to improve growth and yield for chemical bioproduction [12, 40]. Another pair of genes that were mutated several times along parallel experiments is *mrdB* (three times) and *mrdA* (one time). MrdB is annotated as rod shape-determining protein RodA [41, 42] and the related *mrdA* is annotated as penicillin-binding protein 2 [43]. All of the mutations found in the *mrdB* and *mrdA* genes were SNPs changing the properties of the protein through single amino acid changes (see Table 3). The impact of each SNP on the protein functionality or activity was not immediately clear based on the known structural data for these genes. Mutations in the *mrdA* and *mrdB* genes were reported in a previous temperature tolerance ALE experiment [44]. Mutations in *mrd* loci encoding the elongasome such as *mrdA* and *mrdB* have been shown to increase the levels of the growth-rate-regulating molecule (p)ppGpp, which potentially can lead to carbon metabolism modulation [45]. Additionally, a causal mutation in the cell shape determining gene *mrdA* was identified in ALE for developing osmotolerant *E. coli* strains [46].

Validation of mutational causality by reverse engineering

To examine the causality of key mutations identified in this study, growth screens were performed for relevant single and double mutant strains of K-12 MG1655. Such mutant strains (see methods, validation of mutation causality) were either isolated directly from the current study (i.e., endpoint clones containing the csc construct and mutations) or generated from previously constructed mutant strains [22] which were subsequently engineered to also contain the *csc* gene cassette (see Table 4). Growth screens were performed side by side on M9 medium supplemented with either sucrose or glucose (20 g/L). Genes investigated for potential causality for improved fitness, i.e., higher growth rate, were *rpoC*, *rpoB* and the intergenic deletion Δ82 bp between *pyrE*/*rph*. To put these results in context, ALE-derived mutations in these exact genes were validated previously to confer fitness advantages in various substrate environments, and in particular when grown on glucose [22, 32, 33]. Figure 5 provides comparative fitness levels for each adaptive mutation on either sucrose or glucose.

The comparative analysis of reverse engineered strains showed that mutated genes identified in this ALE study conferred a fitness advantage over the starting strain and some were additive in their impact on growth rate. Interestingly, the fitness advantage of all tested mutations over the starting strain was observed when grown on M9 with either sucrose or glucose as carbon sources, and the fitness level increases for a given mutated strain were roughly similar when grown on sucrose or glucose. For the mutations examined, the RNA polymerase mutations have a higher increase in growth rate over the causal

Table 4 Source of key mutations validated for causality in isolated strains from the current study or constructed previously from Lacroix et al. [22]

Genetic region	Mutation	Source
K-12 MG1655	N/A	N/A
pyrE-rph	Δ82 bp deletion	LaCroix et al. [22]
rpoB	E672K (GAA→AAA)	LaCroix et al. [22]
rpoC	R1075C (CGT→TGT)	MGcscBKAp A1 F138 I1 MGcscBKAp A2 F140 I1
	R978C (CGT→TGT)	MGcscBKAp A3 F139 I1
	T1045P (ACC→CCC)	MGcscBKA A3 F140 I1 with additional mutations of (*tdcG*, *baeS*)
pyrE-rph + *rpoB*	Δ82 bp + E672K (GAA→AAA)	LaCroix et al. [22]
pyrE-rph + *rpoB*	Δ82 bp + S621F (TCC→TTC)	MGcscBKA A1 F136 I1

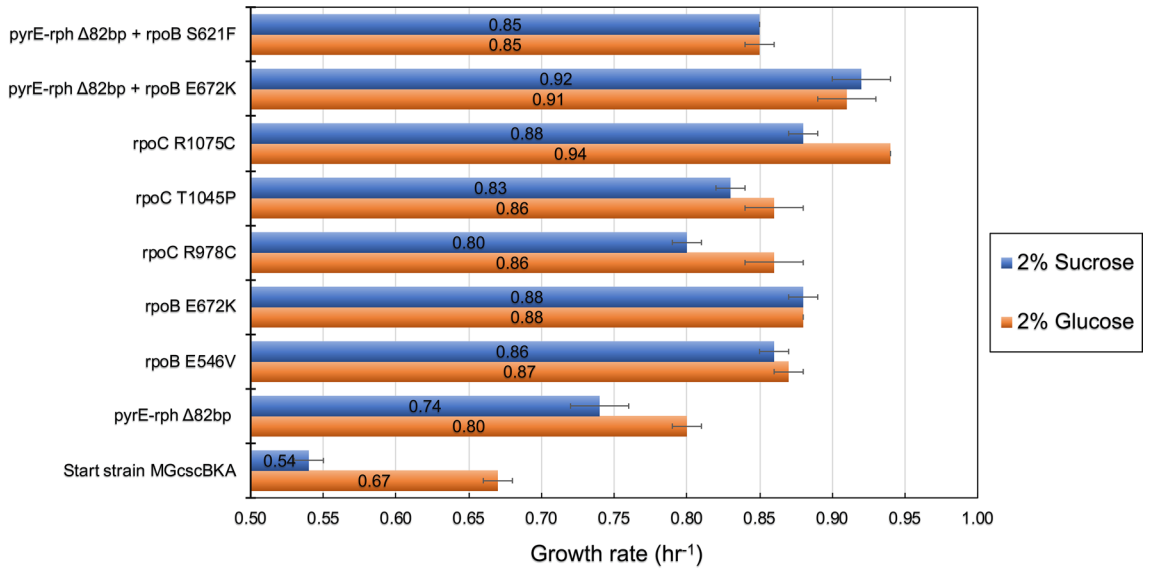


Fig. 5 Causal mutation analysis for fitness improvement on sucrose and glucose for K-12 strains. Shown is a bar graph representing growth rate of identified causal mutations from the current study (ALE on sucrose) and LaCroix et al. [22] (ALE on glucose). The error bars are calculated from two biological replicates (n = 2). All *rpoC* mutations originated from this study, as well as the *rpoB* S621F with the *pyrE-rph* deletion mutation (i.e., they were not reverse engineered). The remaining mutations were from glucose ALE control experiment (GLUALE) with chromosomally integrated csc gene cassette as described in the methods

mutation affecting nucleotide biosynthesis (*pyrE*). Furthermore, analysis showed that when multiple mutations were present in a strain, such double mutant strains conferred a higher fitness advantage over single mutants with the identical mutations. Specifically, a strain with a mutation in the intragenic region between *pyrE*/*rph* genes and in *rpoB* (E672K) had higher fitness than strains with either single mutations. This pattern of co-occurrence was observed when evolving K-12 MG1655 on either glucose [22] or sucrose (endpoint isolate derived from MGcscBKA ALE#1 population from the current study).

Benchmarking of evolved strains performance on sugarcane juice

To further evaluate the ALE-derived clones for sucrose utilization, each endpoint clone was characterized on a raw feedstock, sugarcane juice (SCJ) medium. A M9 medium base was utilized with SCJ to test both evolved and starting strains (Table 2 and Additional file 2: Figs. S1–S3). Raw SCJ mainly consists of three carbon sources; sucrose with the highest fraction, and glucose and fructose in small fractions [47]. The sucrose uptake rates (SURs) and acetate production rates (APRs) for all of the

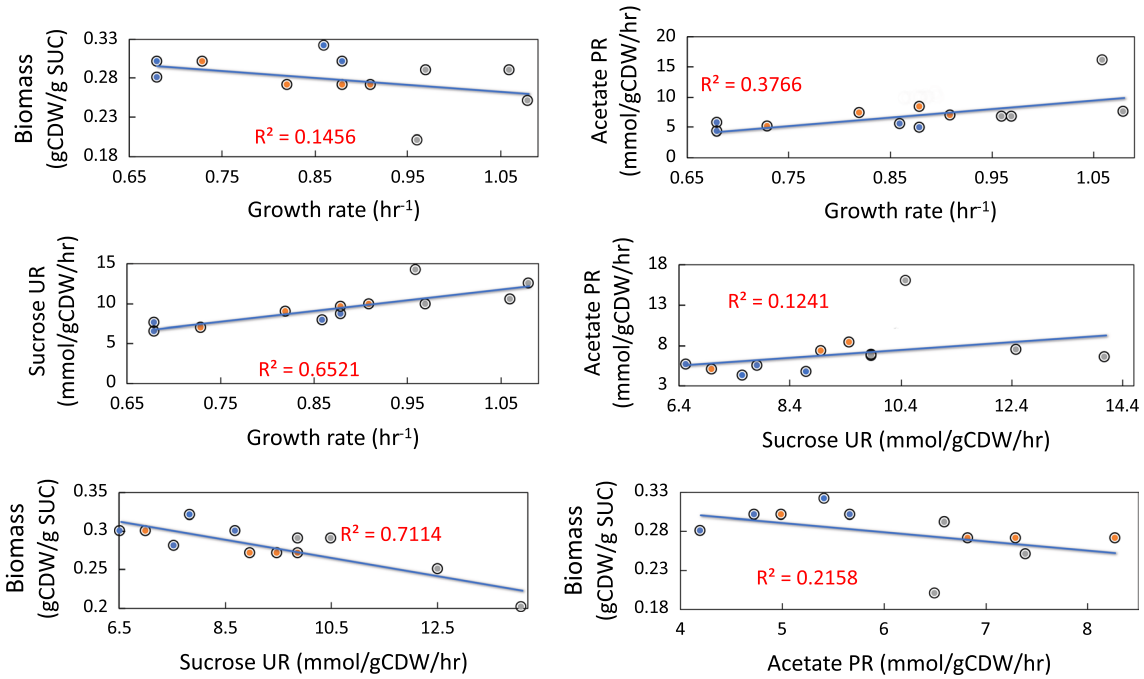


Fig. 6 Pairwise comparisons of the phenotypic properties of the evolved clones on sugarcane juice M9 minimal medium (SCJ medium). Blue filled circles correspond to MGcscBKA clones, orange filled circles to MGcscBKAp clones and grey filled circles to *E. coli* W. Overall, sucrose uptake rate versus biomass yield and growth rate versus sucrose uptake rate were the highly correlated. Calculations can be seen in the methods section

isolated clones were compared to each of the starting strains. Overall, there was an increase in the SUR for the K-12 MG1655 and *E. coli* W endpoint clones. The average increase in SUR was 1.10-±0.05- and 1.35±0.05-fold increase for MGcscBKA and MGcscBKAp isolates, respectively, and a round 1.2±0.2-fold increase for *E. coli* W isolates. Analysis of the MGcscBKAp starting strain showed a slightly lower SUR than that of the MGcscBKA starting strain. This finding was in contrast to the observation that such a mutation in *cscB* leads to a two-fold increase in sucrose uptake rate in a different *E. coli* strain [16]. However, the evolved MGcscBKAp clones demonstrated higher uptake rates than the MGcscBKA evolved clones (Table 2). Thus, it appears that there was an impact from this mutation at faster growth rates displayed by the evolved clones, but not the starting strains. However, reverse engineering of this mutation in the mutated endpoint strains of the MGcscBKA evolved clones would be needed to show this definitively.

As a general observation on the growth profiles of the evolved isolates, the initial amounts of glucose and fructose (approximately, 1–2 mM) were first depleted followed by the consumption of sucrose (Additional file 2: Figs. S1–S3). Interestingly, during the exponential growth phase, fructose and glucose accumulated in the culture broth in some replicates (e.g., *E. coli* W endpoint replicate #1, Additional file 2: Fig. S1) and was then consumed near the end of the aerobic cultivation. It was also observed that acetate was produced and accumulated during the course of the cultivation for all the tested clones except for the *E. coli* W clones (Additional file 2: Figs. S1–S3). In the evolved W clones, acetate accumulated up until approximately 10 h after inoculation, and then it was consumed as the sucrose, glucose, and fructose were depleted.

The physiological profiling of selected endpoint isolate on SCJ were compared to examine whether there were any correlations or observed trends across the different evolved replicate experiments. Although the evolved clones isolated from the endpoints of the ALE experiments mostly showed a similar increase in fitness, the SUR and biomass yields varied more significantly (Table 2). However, the SURs and APRs were most often higher in endpoint clones compared to the starting strains (the exceptions being one clone, MGcscBKA isolate #2, in which SUR was not significant). There was a weak correlation ($r^2=0.12$) between the increase observed in both SUR and APR, Fig. 6. Alternatively, there was a stronger correlation between the increase in SURs and APRs and the increased growth rate with correlations of ($r^2=0.65$) and ($r^2=0.37$), respectively. For the characterized clones, a trade-off between the SURs

and biomass yield at the steady state ($Y_{X/S}$) was observed where higher SURs led to lower biomass yields, i.e. an inverse correlation with an $r^2 = 0.71$. The results implied that there were multiple mutational mechanisms through which a similar increase in the growth rate could be achieved (i.e., alternative optimal solution). The observed variations in SURs, APRs, and biomass yields further showed that under the selection pressure applied via ALE, the populations adapted through different trajectories in order to reach an apparent optimum metabolic state enabling a fast growth phenotype on sugarcane juice.

Discussion

Platform strains that are able to exploit sucrose as abundant cheap carbon source from raw materials sources such as molasses and sugarcane juice can contribute to the increased economic and environmental sustainability of bio-based chemical production. Most of previous attempts to engineer a fast-growing K-12 strain on sucrose, to the same rate as growth on glucose, have not been successful. Thus, the scope of the current study was to generate a platform strains able to utilize sucrose as a carbon source and grow rapidly using adaptive laboratory evolution. Accordingly, the main findings from the current work are: (1) it was possible to effectively generate platform strains with elevated growth rates in a sucrose feedstock environment, which includes validation on raw material sugarcane juice. Moreover, the developed platform strains are suitable as fermentation chassis with an extended capability to also rapidly utilize glucose to the same level as strains evolved to grow solely on glucose [22]; (2) this study establishes a transferable sucrose utilization cassette for use in K-12 strains where sucrose utilization is not native; and (3) the identification of key reproducibly-occurring mutations using multiple parallel replicate ALE experiments. These mutations were validated using knock-in strains and can be used as cell engineering parts in additional strains of interest.

The ALE approach utilized in this work was successful in generating strains with improved fitness over their respective starting strains after approximately 40 days of continuous passaging in the exponential growth phase. The selected clones derived from the evolved populations for the K-12 strains exhibited increased carbon uptake rates for both sucrose and glucose, making them suitable for both carbon sources. Furthermore, the approach utilized in this work to chromosomally integrate the *csc* cassette was successful in generating a genetically stable strain (see Additional file 2: Text), as demonstrated throughout the extent of work in the study. This is in contrast to previously published plasmid expressions that have been shown to lead to complications in industrial production due to requirements for the addition of antibiotics and the burden of maintaining high-copy numbers plasmids [15, 48].

The key mutations identified in this study can be used as engineering parts for strain engineering. Given that a similar set of mutations enables enhanced growth on both glucose and sucrose, it is reasonable to expect that such mutations and strains have growth benefits on similar substrates, or mixes of such substrates, and this represents an avenue forward for generating a more universally applicable platform strain [34, 36, 49]. Previously, the key mutations identified in the RNA polymerase subunits (RNAP) *rpoB* and *rpoC* were found to have large-scale systematic transcriptional changes that influence specific cellular processes and to be responsible for fast growth [34, 36] on minimal medium. The benefits of such mutations in bioprocessing have been demonstrated and establish how strains with such mutations can serve as platform strains for enhanced production phenotypes [50]. Consequently, it appears that the fast-growing phenotype of K-12 on sucrose is more likely more due to a systems level regulatory change rather than changes specific to sucrose utilization.

There are a number of avenues to explore to build off of the results from this study. Such work could include an effort to study the impact of the key mutations individually, or in combination, on the fast growth phenotype. One interesting feature to pursue was that the evolved K-12 strains demonstrated lower SURs on sucrose than the W strain and this could be further examined. It has been shown that *E. coli* W maintains a highly oxidative metabolic state either on glucose or sucrose with low accumulation of overflow metabolites such as acetate [51]. The two *E. coli* strains possess significantly different genome sizes (4.90 Mbp for W vs 4.64 Mbp for K-12) with different numbers of genes, pseudogenes, and mobile elements [10], therefore a comparative genomic study to associate this particular feature was not apparent. A potentially fruitful approach to address this observed difference would likely include growth of the evolved and unevolved *E. coli* W and K-12 strains along with multiple omics assays to better understand the underlying mechanisms, as performed previously [36], in a sucrose utilizing context. Further, such an analysis with an engineered production pathway would also place the evolved phenotypes in a relevant production context.

In summary, ALE has been demonstrated to generate strains with reproducible causal mutations for increased sucrose utilization for *E. coli* under a strict growth rate selection pressure. Whole genome re-sequencing revealed

the underlying causal mutations of the evolved strains. Thus, the ALE derived clones and the beneficial mutations represent a promising platform for developing a sucrose-based bioproduction chassis and can provide a good starting point to develop sucrose utilization in other industrially relevant *E. coli* strains.

Additional files

Additional file 1: Data file S1. All mutation list of ALE resequenced clones and populations.

Additional file 2: Text file S1. Full strains design description and different genetic manipulation methods.

Abbreviations

ALE: adaptive laboratory evolution; SUC: sucrose; SUR: sucrose uptake rate; APR: acetate production rate; GLUALE: glucose adaptive laboratory evolution; RNAP: RNA polymerase; AAR: amino acids residues.

Acknowledgements

We would like to thank Mohammad Radi, Patrick Phaneuf and Anna Koza for their help in the project. Funding for the project was provided by The Novo Nordisk Foundation Grant Numbers NNF14OC0011269 and NNF10CC1016517. We greatly appreciate the help of Tim Mies (Energy Biosciences Institute, University of Illinois for collecting and storing the cane juice utilized in this study. This part of the research was supported by the Energy Biosciences Institute (project OO2J14).

Authors' contributions

Conceived and designed the experiments: ETM, AMF. Performed most of the experiments; ETM. JL constructed the pOSIP-csc plasmids and JL, ETM used it for chromosomal integration of the *csc* cassette. Contributed materials, i.e. Sugarcane Juice: IC and RIM. Wrote the paper: ETM and AMF. HM, JL, MJH, ATN, IC and RIM; reviewed manuscript and helped in discussing the results. All authors read and approved the final manuscript.

Funding

Funding for the project was provided by The Novo Nordisk Foundation Grant Numbers NNF14OC0011269 and NNF10CC1016517.

Availability of data and materials

All data generated or analyzed during this study are included in this published article [and its additional files].

Ethics approval and consent to participate

Not applicable.

Consent of publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Building 220, Kemitorvet, Lyngby 2800 Kgs, Denmark. [2] Department of Animal Sciences, Institute for Genomic Biology and Energy Biosciences Institute, University of Illinois, Urbana, IL 61801, USA. [3] Department of Bioengineering, University of California, 9500 Gilman Drive La Jolla, San Diego, CA 92093, USA.

Received: 3 March 2019 Accepted: 22 June 2019
Published online: 29 June 2019

References

1. Bevan MW, Franssen MCR. Investing in green and white biotech. Nat Biotechnol. 2006;24:765. https://doi.org/10.1038/nbt0706-765.
2. Koutinas AA, Wang R, Webb C. Evaluation of wheat as generic feedstock for chemical production. Ind Crops Prod. 2004;20:75–88.
3. Renouf MAA, Wegener MKK, Nielsen LKK. An environmental life cycle assessment comparing Australian sugarcane with US corn and UK sugar beet as producers of sugars for fermentation. Biomass Bioenerg. 2008;32:1144–55.
4. Fadel M, Keera AA, Mouafi FE, Kahil T. High level ethanol from sugar cane molasses by a new thermotolerant *Saccharomyces cerevisiae* strain in industrial scale. Biotechnol Res Int. 2013;2013:253286.
5. Jung MY, Jung HM, Lee J, Oh MK. Alleviation of carbon catabolite repression in *Enterobacter aerogenes* for efficient utilization of sugarcane molasses for 2,3-butanediol production. Biotechnol Biofuels. 2015;8:106.
6. Monk JMM, Koza A, Campodonico MAA, Machado D, Seoane JMM, Palsson BOO, et al. Multi-omics quantification of species variation of *Escherichia coli* links molecular features with strain phenotypes. Cell Syst. 2016;3(238–251):e12.
7. Baeshen MN, Al-Hejin AM, Bora RS, Ahmed MMM, Ramadan HAI, Saini KS, et al. Production of biopharmaceuticals in *E. coli*: current scenario and future perspectives. J Microbiol Biotechnol. 2015;25:953–62.
8. Idalia V-MN, Bernardo F. *Escherichia coli* as a model organism and its application in biotechnology. Recent Adv Physiol Pathog Biotechnol Appl. 2017. https://doi.org/10.5772/67306.
9. Bergey DH, David H, Holt JG. Bergey's manual of determinative bacteriology. Medical; 1994. https://books.google.se/books/about/Bergey_s_Manual_of_Determinative_Bacteri.html?id=jtMLzaa5ONcC&redir_esc=y. Accessed 4 Jun 2019.
10. Archer CT, Kim JF, Jeong H, Park JH, Vickers CE, Lee SY, et al. The genome sequence of *E. coli* W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. BMC Genomics. 2011;12:9. https://doi.org/10.1186/1471-2164-12-9.
11. Bockmann J, Heuel H, Lengeler JW. Characterization of a chromosomally encoded, non-PTS metabolic pathway for sucrose utilization in *Escherichia coli* EC3132. Mol Gen Genet. 1992;235:22–32.
12. Sabri S, Nielsen LK, Vickers CE. Molecular control of sucrose utilization in Escherichia coli W, an efficient sucrose-utilizing strain. Appl Environ Microbiol. 2013;79:478–87.
13. Cheng L-CC, Hor L-II, Wu J-YY, Chen T-LL. Effect of specific growth rate on the production of a recombinant nuclease by *Escherichia coli*. Biochem Eng J. 2003;14:101–7.
14. Zawada J, Swartz J. Maintaining rapid growth in moderate-density *Escherichia coli* fermentations. Biotechnol Bioeng. 2005;89:407–15. https://doi.org/10.1002/bit.20369.
15. Diaz Ricci JC, Hernández ME. Plasmid effects on *Escherichia coli* metabolism. Crit Rev Biotechnol. 2000. https://doi.org/10.1080/07388550008984167.
16. Jahreis K, Bentler L, Bockmann J, Hans S, Meyer A, Siepelmeyer J, et al. Adaptation of sucrose metabolism in the *Escherichia coli* wild-type strain EC3132. J Bacteriol. 2002;184:5307–16.
17. Lee JW, Choi S, Park JH, Vickers CE, Nielsen LK, Lee SY. Development of sucrose-utilizing *Escherichia coli* K-12 strain by cloning β-fructofuranosidases and its application for L-threonine production. Appl Microbiol Biotechnol. 2010;88:905–13.
18. Shukla VB, Zhou S, Yomano LP, Shanmugam KT, Preston JF, Ingram LO. Production of L(−)-lactate from sucrose and molasses. Biotechnol Lett. 2004;26:689–93. https://doi.org/10.1023/B:BILE.0000024088.36803.4e.
19. Tsunekawa H, Azuma S, Okabe M, Okamoto R, Aiba S. Acquisition of a sucrose utilization system in *Escherichia coli* K-12 derivatives and its application to industry. Appl Environ Microbiol. 1992;58:2081–8.
20. Bruschi M, Boyes SJ, Sugiarto H, Nielsen LK, Vickers CE. A transferable sucrose utilization approach for non-sucrose-utilizing *Escherichia coli* strains. Biotechnol Adv. 2012;30:1001–10.
21. Lee SY, Kim HU. Systems strategies for developing industrial microbial strains. Nat Biotechnol. 2015;33:1061–72.
22. LaCroix RA, Sandberg TE, O'Brien EJ, Utrilla J, Ebrahim A, Guzman GI, et al. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of *Escherichia coli* K-12 MG1655 on glucose minimal medium. Appl Environ Microbiol. 2015;81:17–30. https://doi.org/10.1128/AEM.02246-14.

23. Sandberg TE, Pedersen M, LaCroix RA, Ebrahim A, Bonde M, Herrgard MJ, et al. Evolution of *Escherichia coli* to 42 C and subsequent genetic engineering reveals adaptive mechanisms and novel mutations. Mol Biol Evol. 2014;31:2647–62.
24. St-Pierre F, Cui L, Priest DG, Endy D, Dodd IB, Shearwin KE. One-step cloning and chromosomal integration of DNA. ACS Synth Biol. 2013;2:537–41. https://doi.org/10.1021/sb400021j.
25. Cui L, Shearwin KE. Clonetegration using OSIP plasmids: one-step DNA assembly and site-specific genomic integration in bacteria. 2017. p. 139–55. https://doi.org/10.1007/978-1-4939-6343-0_11.
26. Phaneuf PV, Gosting D, Palsson BO, Feist AM. ALEdb 1.0: a database of mutations from adaptive laboratory evolution. bioRxiv. 2018;320747. https://doi.org/10.1101/320747v1.full.
27. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol. 2014;1151:165–88.
28. LaCroix RA, Palsson BO, Feist AM. A model for designing adaptive laboratory evolution experiments. Appl Environ Microbiol. 2017. https://doi.org/10.1128/aem.03115-16.
29. Lee DH, Feist AM, Barrett CL, Palsson B. Cumulative number of cell divisions as a meaningful timescale for adaptive laboratory evolution of *Escherichia coli*. PLoS ONE. 2011;6:e26172. https://doi.org/10.1371/journal.pone.0026172.
30. Ross W, Gosink K, Salomon J, Igarashi K, Zou C, Ishihama A, et al. A third recognition element in bacterial promoters: DNA binding by the subunit of RNA polymerase. Science. 1993;262:1407–13.
31. Simpson RB. The molecular topography of rna polymerase-promoter interaction. Cell. 1979;18:277–85.
32. Sandberg TE, Lloyd CJ, Palsson BO, Feist AM. Laboratory evolution to alternating substrate environments yields distinct phenotypic and genetic adaptive strategies. Appl Environ Microbiol. 2017. https://doi.org/10.1128/aem.00410-17.
33. Long CP, Gonzalez JE, Feist AM, Palsson BO, Antoniewicz MR. Fast growth phenotype of *E. coli* K-12 from adaptive laboratory evolution does not require intracellular flux rewiring. Metab Eng. 2017;44:100–7. https://doi.org/10.1016/j.ymben.2017.09.012.
34. Conrad TM, Frazier M, Joyce AR, Cho B-K, Knight EM, Lewis NE, et al. RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. Proc Natl Acad Sci USA. 2010;107:20500–5.
35. Mohamed ET, Wang S, Lennen RM, Herrgård MJ, Simmons BA, Singer SW, et al. Generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution. Microb Cell Fact. 2017;16:1–15. https://doi.org/10.1186/s12934-017-0819-1.
36. Utrilla J, O'Brien EJ, Chen K, McCloskey D, Cheung J, Wang H, et al. Global rebalancing of cellular resources by pleiotropic point mutations illustrates a multi-scale mechanism of adaptive evolution. Cell Syst. 2016;2:260–71. https://doi.org/10.1016/j.cels.2016.04.003.
37. Jensen KF. The *Escherichia coli* K-12 "wild types" W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. J Bacteriol. 1993;175:3401–7.
38. Norregaard-Madsen M, McFall E, Valentin-Hansen P. Organization and transcriptional regulation of the *Escherichia coli* K-12 D-serine tolerance locus. J Bacteriol. 1995;177:6456–61.
39. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44:D279–85.
40. Arifin Y, Sabri S, Sugiarto H, Krömer JO, Vickers CE, Nielsen LK. Deletion of cscR in *Escherichia coli* W improves growth and poly-3-hydroxybutyrate (PHB) production from sucrose in fed batch culture. J Biotechnol. 2011;156:275–8.
41. Meeske AJ, Riley EP, Robins WP, Uehara T, Mekalanos JJ, Kahne D, et al. SEDS proteins are a widespread family of bacterial cell wall polymerases. Nature. 2016;537:634–8.
42. Cho H, Wivagg CN, Kapoor M, Barry Z, Rohs PDA, Suh H, et al. Bacterial cell wall biogenesis is mediated by SEDS and PBP polymerase families functioning semi-autonomously. Nat Microbiol. 2016;1:16172.
43. Ishino F, Park W, Tomioka S, Tamaki S, Takase I, Kunugita K, et al. Peptidoglycan synthetic activities in membranes of *Escherichia coli* caused by overproduction of penicillin-binding protein 2 and RodA protein. J Biol Chem. 1986;261:7024–31.
44. Deatherage DE, Kepner JL, Bennett AF, Lenski RE, Barrick JE. Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures. Proc Natl Acad Sci. 2017;114:E1904–12.
45. Sperber AM, Herman JK. Metabolism shapes the cell. J Bacteriol. 2017;199:1–14.
46. Winkler JD, Garcia C, Olson M, Callaway E, Kao KC. Evolved osmotolerant escherichia coli mutants frequently exhibit defective *N*-acetylglucosamine catabolism and point mutations in cell shape-regulating protein MreB. Appl Environ Microbiol. 2014;80:3729–40.
47. Godshall MA, Legendre BL. SUGAR|Sugarcane. Encycl Food Sci Nutr. 2003;56:45–51.
48. Wang Z, Xiang L, Shao J, Wegrzyn A, Wegrzyn G. Effects of the presence of ColE1 plasmid DNA in *Escherichia coli* on the host cell metabolism. Microb Cell Fact. 2006;5:1–18.
49. Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, et al. Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol. 2009;10:R118.
50. Rugbjerg P, Feist AM, Sommer MOA. Enhanced metabolite productivity of *Escherichia coli* adapted to glucose m9 minimal medium. Front Bioeng Biotechnol. 2018;6:166. https://doi.org/10.3389/fbioe.2018.00166/full.
51. Arifin Y, Archer C, Lim SA, Quek LEE, Sugiarto H, Marcellin E, et al. *Escherichia coli* W shows fast, highly oxidative sucrose metabolism and low acetate formation. Appl Microbiol. 2014;98:9033–44. https://doi.org/10.1007/s00253-014-5956-4.

Publisher's Note

Paper 2: Generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution

Mohamed *et al. Microb Cell Fact (2017) 16:204*
DOI 10.1186/s12934-017-0819-1

Microbial Cell Factories

RESEARCH

Open Access

Generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution

Elsayed T. Mohamed[1], Shizeng Wang[3,4,5], Rebecca M. Lennen[1], Markus J. Herrgård[1], Blake A. Simmons[3,4], Steven W. Singer[3,4] and Adam M. Feist[1,2*]

Abstract

Background: There is a need to replace petroleum-derived with sustainable feedstocks for chemical production. Certain biomass feedstocks can meet this need as abundant, diverse, and renewable resources. Specific ionic liquids (ILs) can play a role in this process as promising candidates for chemical pretreatment and deconstruction of plant-based biomass feedstocks as they efficiently release carbohydrates which can be fermented. However, the most efficient pretreatment ILs are highly toxic to biological systems, such as microbial fermentations, and hinder subsequent bioprocessing of fermentative sugars obtained from IL-treated biomass.

Methods: To generate strains capable of tolerating residual ILs present in treated feedstocks, a tolerance adaptive laboratory evolution (TALE) approach was developed and utilized to improve growth of two different *Escherichia coli* strains, DH1 and K-12 MG1655, in the presence of two different ionic liquids, 1-ethyl-3-methylimidazolium acetate ($[C_2C_1Im][OAc]$) and 1-butyl-3-methylimidazolium chloride ($[C_4C_1Im]Cl$). For multiple parallel replicate populations of *E. coli*, cells were repeatedly passed to select for improved fitness over the course of approximately 40 days. Clonal isolates were screened and the best performing isolates were subjected to whole genome sequencing.

Results: The most prevalent mutations in tolerant clones occurred in transport processes related to the functions of *mdtJI*, a multidrug efflux pump, and *yhdP*, an uncharacterized transporter. Additional mutations were enriched in processes such as transcriptional regulation and nucleotide biosynthesis. Finally, the best-performing strains were compared to previously characterized tolerant strains and showed superior performance in tolerance of different IL and media combinations (i.e., cross tolerance) with robust growth at 8.5% (w/v) and detectable growth up to 11.9% (w/v) $[C_2C_1Im][OAc]$.

Conclusion: The generated strains thus represent the best performing platform strains available for bioproduction utilizing IL-treated renewable substrates, and the TALE method was highly successful in overcoming the general issue of substrate toxicity and has great promise for use in tolerance engineering.

Keywords: *Escherichia coli*, Renewable feedstocks, Ionic liquids, Adaptive laboratory evolution

Background

There is a need to replace chemical and fuel production from fossil feedstocks with carbon neutral sources to retain the natural cycle of carbon emission and assimilation. Certain biomass feedstocks can play a major part in this need as they are abundant, diverse, and renewable. These biomass feedstocks include general plant-based materials like energy crops, crop residues, wood, wood residues, and grasses. Most of these materials have intrinsic value alongside with the added possibility of use as biomaterials [1]. Biomass feedstocks, however, possess a low energy density requiring a greater quantity of them to meet market demands [2].

*Correspondence: afeist@ucsd.edu
[2] Department of Bioengineering, University of California, 9500 Gilman Drive La Jolla, San Diego, CA 92093, USA
Full list of author information is available at the end of the article

Therefore, innovative approaches are necessary to make biomass feedstocks viable carbon sources to replace fossil feedstocks.

Lignocellulosic biomass can serve as a carbon neutral and abundant feedstock for bioprocesses [3]. In order to utilize lignocellulosic biomass for biochemical conversion in biorefineries, a pretreatment process is needed to remove the physical and chemical barriers to fully utilize the sugar substrates. The main aim of pretreatment is to increase the accessibility of cellulose, which then can be subjected to enzymatic saccharification to release fermentable sugars. This can be achieved through dissolution of hemicellulose and/or lignin, which coat the surface of cellulose [4]. There are several approaches for pretreatment of lignocellulosic biomass which include physical and chemical methods, but one of the most effective approaches is to directly release monomeric sugars through treatment with Ionic liquids (ILs) [5]. ILs are effective solvents for deconstruction and result in generating sugar feedstocks without significant loss of sugars due to degradation [4, 6, 7].

Despite its efficacy, IL pretreatment has some limitations, as some amounts of ILs remain from the pretreatment and these are often highly toxic to microbes used in downstream fermentation processes. Typically, a deconstruction hydrolysate has around 0.2–5% (w/v) of ionic liquid after the pretreatment [8]. One practice to overcome toxicity is to wash the pretreatment several times mostly with water, but this process adds significant purification costs [6]. Thus, an alternate approach to deal with toxicity by developing microbial platform strains that can tolerate residual ILs is needed. Limited tolerance toward ILs has been previously achieved using a range of techniques including rational design [8–10] and adaptive laboratory evolution (ALE) [11]. The rationally-designed strains generally introduced a non-native efflux pump for ILs which exerts a metabolic burden on the cells and a need for tight expression control. The previous ALE study showed promise for the approach [11], but the scope was limited to one IL, utilized a rich undefined media, and no genetic basis for the improved performance was presented. Nonetheless, this preliminary work revealed an opportunity to apply ALE for IL tolerance and can be used for comparison.

In the present work, the problem of IL toxicity is addressed using a systematic ALE strategy, Tolerance adaptive laboratory evolution (TALE), to generate *Escherichia coli* (*E. coli*) strains that were highly tolerant to the presence of ILs. The TALE method differs from previous efforts in that dynamic control is used to increase the amount of stress applied to cells to keep a strong selection pressure without crashing cultures due to overstressed growth conditions. The TALE approach for IL tolerance employed in the present study has two major advantages compared to manual ALE work (e.g., [11]). First, the TALE approach significantly improved fitness and final cell density in higher IL concentrations than the manual ALE approach. Additionally, the IL cross-tolerance phenotype exhibited by the best performing strains can expand the application of TALE-derived strains. Finally, these results were obtained over a significantly shorter time frame (40 vs 90 days) using an automated platform for performing TALE (details of the improvement are provided below).

In this study, two biotechnologically-relevant strains of *E. coli* (K-12 MG1655 and DH1) were exposed to increasing concentrations of two ILs; 1-butyl-3-methylimidazolium chloride and 1-ethyl-3-methylimidazolium acetate. Both of these targeted ILs are promising solvents for biomass pretreatment and were considered as a good candidates for IL-pretreated biomass [5, 7]. The exposure was performed over repeated exponential batch growth in parallel biological replicates. The evolved populations were screened and individual isolates were re-sequenced to identify key causal mutations. Selected isolates were compared against rationally-designed strains previously demonstrated to possess IL tolerance [8, 12]. The best performing strains showed markedly improved tolerance toward higher concentrations of ILs over rationally designed strains. The key mutations identified in this study provide a linkage between the IL tolerance phenotype and genotype.

Methods

Strains, reagents and equipment

Two *E. coli* strains were utilized: DH1 (ATCC 33849) and K-12 MG1655 (ATCC 47076). Two ionic liquids were utilized. 1-Butyl-3-methylimidazolium chloride ($[C_4C_1Im]Cl$) was purchased from Sigma-Aldrich (Basionics ST 70, BASF), and 1-ethyl-3-methylimidazolium acetate, ($[C_2C_1Im][OAc]$), was purchased from IOLITEC ionic Liquids Technologies GmbH (Heilbronn, Germany). Chemicals and components of the medium used for selecting the best performing strains were purchased from Sigma-Aldrich (St. Louis, USA) or VWR (West Chester, USA) unless otherwise noted.

M9 glucose medium contained 2 g/L glucose, 1× M9 salts, 2 mM $MgSO_4$, 100 µM $CaCl_2$ and 1× trace elements and Wolfe's vitamin solution. Composition of 10× M9 salts solution consisted of 68 g/L Na_2HPO_4 anhydrous, 30 g/L KH_2PO_4, 5 g/L NaCl, and 10 g/L NH_4Cl dissolved in Milli-Q filtered water and autoclaved. M9 trace elements was a 2000× solution containing of 3.0 g/L $FeSO_4{\cdot}7H_2O$, 4.5 g/L $ZnSO_4{\cdot}7H_2O$, 0.3 g/L $CoCl_2{\cdot}6H_2O$, 0.4 g/L $Na_2MoO_4{\cdot}2H_2O$, 4.5 g/L $CaCl_2{\cdot}H_2O$, 0.2 g/L $CuSO_4{\cdot}2H_2O$, 1.0 g/L H_3BO_3, 15 g/L

disodium ethylene-diamine-tetra-acetate, 0.1 g/L KI, 0.7 g/L $MnCl_2{\cdot}4H_2O$ and concentrated HCl dissolved in Milli-Q filtered water and sterile filtered. The final concentrations of the vitamin mix and trace elements in the M9 medium were 1×.

Screening for tolerance in wild type strains

The two *E. coli* strains, DH1 and K-12 MG1655, were initially screened for their tolerance towards different concentrations of each IL in order to choose the starting concentration where the growth rate and final optical density were higher. A description of tolerance screening and tolerance phenotype in wild type strains (Additional file 2: Table S1). Cells from an overnight culture in LB medium were inoculated into cylindrical tubes containing 15 mL M9 glucose supplemented with varying concentrations of each ionic liquid. Inoculated tubes were temperature-controlled at 37 °C and fully aerated. Growth rates and final optical density were determined from 600 nm wavelength (OD_{600}) measurements on a sunrise plate reader (Tecan, Männedorf, Switzerland).

Adaptive laboratory evolution of IL tolerance

The bacterial cells were adaptively evolved under batch fermentation in M9 glucose supplemented with the initial ionic liquid concentration listed in Table 1, with increasing concentration of ILs applied over the course of the ALEs. Cells were serially passaged during exponential growth for approximately 40 days using an automated liquid-handler platform [13]. Pre-cultures for inoculating the starting culture were grown in M9 glucose and 150 µL of each pre-culture was used to inoculate each independent replicate with a working volume of 15 mL. Cells were cultured at 37 °C. OD_{600} was measured at a time determined algorithmically and once OD_{600} reached approximately OD_{600} 0.3, 150 µL was passed into a new tube with a fresh media containing ILs and a total working volume of 15 mL (i.e., a 1:100 ratio). The commonly experienced exponential growth phase was from time of inoculation to approximately OD_{600} 0.3 and the maximum final OD_{600} was approximately 0.4, thus the cells were passed during the exponential phase. The OD_{600} was measured by a Sunrise Plate Reader (Tecan Inc., Switzerland) and the common ratio between the plate reader OD_{600} and a benchtop spectrophotometer with a 1 cm path length is 4.2. Growth rates were determined by calculating the slope of the semi-log plot of log OD versus time using linear regression with the Polyfit function in MATLAB (The Mathworks Inc., Natick, Massachusetts). When increased growth rate was achieved after a defined period

Table 1 Growth phenotypes for *E. coli* K-12 MG1655 and DH1 evolved populations endpoints on ILs ([C_4C_1Im]Cl and [C_2C_1Im][OAc])

Ionic liquid	Strains	Replicate #	Starting conc. %	Initial growth rate (h^{-1})[a]	Average end conc. %	Final growth rate for end conc. (h^{-1})[a]	Change in IL conc. (%)	Total number of flasks
[C_4C_1Im]Cl	MG1655	ALE #1	1.5	0.3	6.2	0.2 ± 0.02	4.7	62
	MG1655	ALE #2	1.5	0.2	6.2	0.1 ± 0.03	4.7	67
	MG1655	ALE #3	1.5	0.3	4.9	0.3 ± 0.12	3.4	63
	MG1655	ALE #4	1.5	0.2	5.6	0.1 ± 0.05	4.1	84
					5.7 ± 0.6	0.2		
	DH1	ALE #5	1.5	0.2	4.8	0.2 ± 0.05	3.3	61
	DH1	ALE #6	1.5	0.2	5.6	0.3 ± 0.08	4.1	72
	DH1	ALE #7	1.5	0.2	5.6	0.1 ± 0.01	4.1	79
	DH1	ALE #8	1.5	0.2	4.2	0.2 ± 0.02	2.7	50
					5.0 ± 0.6	0.2		
[C_2C_1Im][OAc]	MG1655	ALE #9	2	0.2	5.9	0.1 ± 0.00	3.9	86
	MG1655	ALE #10	2	0.2	5.9	0.1 ± 0.04	3.9	87
	MG1655	ALE #11	2	0.2	6.5	0.2 ± 0.00	4.5	91
	MG1655	ALE #12	2	0.2	5.9	0.1 ± 0.01	3.9	92
					6.1 ± 0.3	0.1		
	DH1	ALE #13	1	0.1	4.5	0.1 ± 0.06	3.5	77
	DH1	ALE #14	1	0.1	4.5	0.1 ± 0.02	3.5	92
	DH1	ALE #15	1	0.2	4.5	0.2 ± 0.01	3.5	79
	DH1	ALE #16	1	0.2	5.2	0.2 ± 0.02	4.2	88
					4.6 ± 0.3	0.2		

[a] Initial and final growth rates were calculated for the whole population from the first and last 3 flasks of each corresponding population endpoints, respectively

of time at a particular concentration, the ionic liquid concentration was increased. This process was repeated until a significant increase in tolerance was achieved. Periodically, samples were frozen in a 25% v/v glycerol solution and stored at − 80 °C for further use.

Primary screening

Evolved isolates were screened for growth properties (growth rate, lag time, and final OD_{600}) on selected concentrations (Table 1) of $[C_4C_1Im]Cl$ and $[C_2C_1Im][OAc]$ using a Growth Profiler (EnzyScreen BV, Leiden, Netherlands). Populations from endpoint evolutions were plated on LB agar plates and ten individual colonies derived from each population were screened at the maximum concentration for which robust growth rates were achieved during the evolution. Colonies from wild type starting strains, *E. coli* K-12 MG1655 and DH1, were used as a control for the primary screening. Selected isolates were inoculated into 500 µL M9 glucose medium in deep well plates and incubated in a plate shaker at 37 °C and 300 rpm shaking. Later, cells were diluted 10× in M9 glucose medium, from which 30 µL was transferred to clear-bottom 96 half-deepwell plates (EnzyScreen BV, Leiden, Netherlands) containing M9 glucose medium with either of the two ionic liquids, such that the final concentration was equal to the concentrations listed in Table 1. Cryogenic stocks of the pre-culture plates were stored in 96-well plates. The half-deepwell plates were incubated at 37 °C with 225 rpm shaking in the Growth Profiler, with scans recorded at 15 min intervals. Green pixel (G) values extracted from the 1 mm diameter circular areas in the center of each well in the images were converted to OD_{600} values using a calibration between OD_{600} (1 cm path-length) and G values. The resulting growth curves for each isolate, Additional file 2: Figure S1, were inspected for those exhibiting robust growth or unique growth profiles such as exhibiting reduced lag-times, increased final densities, and increased in the apparent growth rates. Ten isolates from each population were grouped according to their similarities between these parameters.

Secondary screening of TALE isolates

Three individual isolates chosen from primary screening from each population underwent secondary screening, where biological replicates were analyzed. The IL concentration was lowered to the average concentration used in the primary screen for all clones. These adjusted IL concentrations for both *E. coli* K-12 MG1655 and DH1 are listed in Additional file 2: Tables S2 and S3, respectively. Selected isolates from the primary screening were steaked out on LB agar from the cryogenic stock plates stored for primary screening. Three individual colonies from each isolate were inoculated as biological replicates into 96-well deepwell plates containing 500 µL M9 glucose and grown overnight. The next day, cryogenic stocks were prepared as described for primary screening. Each well from the overnight culture were inoculated to a low OD_{600} (1:100 dilution) in M9 glucose medium with the specified IL concentration, and growth was monitored until stationary phase was reached. Growth rates were calculated as described previously, and the average values of the three cultures were determined.

Re-sequencing of improved IL tolerance clones

Sequencing was performed on an Illumina NextSeq (Model 550) Sequencer (San Diego, CA). A total of 45 isolates were re-sequenced. Selected colonies were isolated on LB agar plates, genomic DNA was extracted using PureLink® Genomic DNA Kits (Invitrogen, CA). The quality of extracted DNA was assessed with UV absorbance ratios using a nano drop. Concentration of DNA was quantified using Qubit ds-DNA high sensitivity assay. Paired-end resequencing libraries were generated using a 300 cycle (150 bp × 2) kit from Illumina (San Diego, CA) with loading concentration on Nextseq 1.2 pico-Molar with 1% PhiX spike (Illumina, San Diego, CA) of input DNA total. Re-sequencing data were analyzed using a customized script based on the Breseq version 0.30.1 [14] to map sequence reads and identify mutations relative to the reference strain. The average coverage for each isolate was typically over 400 (a relatively high coverage for clonal sequencing). The genomes of the evolved strains were sequenced and mapped to the genome of the parent strains (NCBI Accession Numbers NC_000913.3 and NC_017625.1) to examine mutations.

Criteria for choosing the best-performing clones from the secondary screen

Additional file 2: Figure S2 summarizes the method used to choose representative clones from each genotypic-clustered set. These representative clones were being used to screen enhanced performance for each clone. For each genotypic-clustered set, if the physiology was the same with a %RSD (relative standard deviation) $\leq$ 20% variability in growth rate and final OD, the selection was made based on the clone with the fastest growth rate and highest final OD with least variability. Alternatively, clones with the fastest growth rate and highest final OD were selected.

Comparing TALE best-performing clones to rational-designed strains

The medium used was a modified M9 glucose containing 4 g/L glucose. The appropriate amounts of antibiotics (100 mg/L carbenicillin and 50 mg/L kanamycin) were

added when needed. $[C_2C_1Im][OAc]$ (BASF, Germany) was added to the medium as indicated. Seed cultures in LB medium were grown overnight and diluted 1:50 into M9 glucose medium (modified to contain 4 g/L glucose, with 100 mg/L carbenicillin and 50 mg/L kanamycin added when needed) containing varying concentrations of $[C_2C_1Im][OAc]$ or $[C_2C_1Im]Cl$ (BASF, Germany). Cultures were grown in 96 well plates (BD Falcon) on a Tecan F-200 Pro microtiter plate reader (Maennedorf, Switzerland) or in 250 mL flat-bottom flasks at 200 rpm and 37 °C. The absorbance of samples from 250 mL flat-bottom flasks was measured at 600 nm using a Spectramax M2 spectrophotometer (Molecular Devices, USA). Initially, selected clones were screened in $[C_2C_1Im][OAc]$ to determine levels of tolerance to a IL not used for the ALE process (i.e., cross-tolerance), then best-performing clones were used for the comparison in $[C_2C_1Im][OAc]$ and $[C_2C_1Im]Cl$ in either M9 or LB medium to understand performance in both a minimal defined and rich undefined media. The parent strain *E. coli* K-12 MG1655 was used as a control. The IL tolerant strains JBEI-10101 [8] and JBEI-13314 [12] were also tested for comparison.

Results

A tolerance adaptive laboratory evolution (TALE) experiment was utilized to generate strains which could tolerate toxic concentrations of similarly close ionic liquids (ILs) and identify mutations which likely confer a fitness advantage under more economically advantageous bioprocessing conditions, i.e., pretreated biomass solution containing some ILs. Two different *E. coli* strains were chosen for the study (K-12 MG1655 and DH1) as well as two types of ILs; 1-butyl-3-methylimidazolium chloride ($[C_4C_1Im]Cl$) and 1-ethyl-3-methylimidazolium acetate ($[C_2C_1Im][OAc]$). The *E. coli* strains were chosen as they are often used in bioprocessing applications [15, 16] and as the adaptive responses of K-12 MG1655 toward minimal medium growth is known [17]. IL pretreatment for biomass deconstruction has been demonstrated in several studies as a promising approach to solubilize cellulosic polysaccharides, thereby increasing the enzymatic turnover of saccharification, and also reducing the formation of inhibitory by-products [6, 18, 19].

Description of fitness changes during the TALE experiment

The process of TALE was successful in generating strains with increased tolerance to ILs. Four replicate populations of each *E. coli* strain were evolved on each of the ILs. The IL concentration was increased under continuous exponential batch growth over the course of the experiment when an observed fitness over a threshold was achieved (i.e., a growth rate of $\geq$ 0.15/h). During this period of time, each of the populations underwent a total number of cumulative cells divisions (CCD) of 1.93×10^{12}, 2.73×10^{12} for *E. coli* K-12 MG1655 and 1.56×10^{12}, 1.61×10^{12} CCD for DH1 populations with $[C_4C_1Im]Cl$ and $[C_2C_1Im][OAc]$, respectively (Additional file 2: Table S4). The use of CCD has previously been shown to be a meaningful measure of evolution time [20]. The observed growth rates during a representative TALE

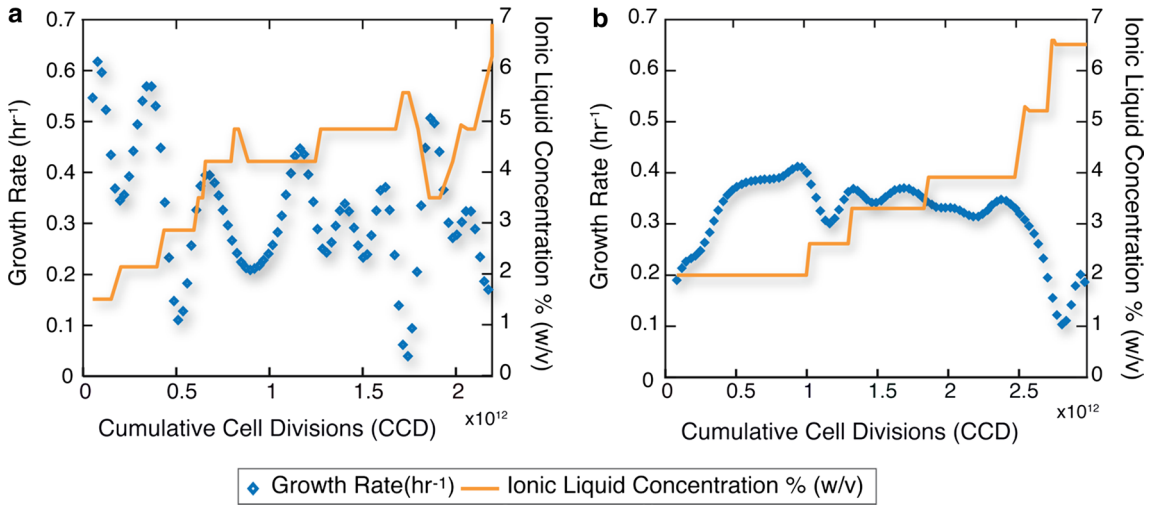


Fig. 1 Plots of population growth rate versus IL concentration over the course of two TALE experiments for increasing IL tolerance. **a** *E. coli* K-12 MG1655 population ALE #1 evolved on glucose minimal medium with ($[C_4C_1Im]Cl$) and, **b** *E. coli* DH1 population ALE #11 evolved on glucose minimal medium with ($[C_2C_1Im][OAc]$). Depicted are fitness trajectories and IL concentration versus cumulative cell divisions (CCD) experienced by the cultures for two out of the total sixteen individual experiments. IL concentration was increased step-wise when the growth rate for the population increased

experiment as well as the initial and final IL concentrations are shown in Fig. 1. Similar plots for the remaining populations are shown in Additional file 2: Figure S3.

The ability of K-12 MG1655 populations to adapt to increasing IL concentrations was superior to DH1 for both ionic liquids, with average final concentrations achieved of 5.7 ± 0.6% (w/v) and 6.1 ± 0.3% (w/v) for MG1655 and 5.0 ± 0.6% (w/v) and 4.6 ± 0.3% (w/v) for DH1, with $[C_4C_1Im]Cl$ and $[C_2C_1Im][OAc]$, respectively (Table 1). During the TALE experiments, population fitnesses fluctuated in response to the concentration of IL added to media (Fig. 1). Additionally, applied IL concentration increases were sometimes too large and resulted in ceased growth. In these instances, the concentration was adjusted back to the previous concentration in order to restore growth, and a smaller step change in concentration was employed. Overall, there were approximately five increases in IL concentration for each experiment using an average step increase of 0.75% (w/v) over the current concentration. Each experiment contained an average of 67 flasks in $[C_4C_1Im]Cl$ and an average of 87 flasks in $[C_2C_1Im][OAc]$. Screening of the evolved populations was subsequently performed to understand the overall tolerance and performance of evolved isolates.

Screening of evolved isolates for improved tolerance

Isolates from evolved populations were screened for improved tolerance to ILs and to help identify causal mutations through genotype–phenotype relationships following resequencing. A primary screen was performed to establish whether selected isolates from each population (10 isolates from each of 16 populations) could grow reproducibly in the average final IL concentration achieved during TALE (Table 1). From this analysis, three isolates from each population were selected for secondary screening and whole genome sequencing based on qualitative differences in the observed primary screen growth rates (Additional file 2: Figure S1). Additionally, isolates were selected that still exhibited tolerance but displayed unique growth phenotypes. Clones from one of the populations of DH1 on $[C_4C_1Im]Cl$ did not grow; therefore, this population was dropped from the analysis (ALE #8). A total of 45 isolates were whole genome resequenced.

Whole genome resequencing and mutation analysis

Whole genome resequencing was used to determine the genetic basis of fitness tolerance phenotypes. Key mutations were determined by comparing all of the clones and identifying genes, or genetic regions (i.e., intergenic regions) which had multiple unique mutations or were mutated across isolates from independent populations. Overall, there were 37 and 53 unique mutations identified for *E. coli* K-12 MG1655 and DH1, respectively. Each isolate had between 1 and 13, or 2 and 12 mutations identified for MG1655 or DH1, respectively. There were three hyper-mutator isolates (1 from MG1655 and 2 from DH1) identified. The MG1655 isolate had 267 mutations, while the DH1 isolates had 39 ± 4 mutations, as compared to the average mutations 5 ± 4 for the non-hypermutating clones from both strains with standard deviation shown between replicates. The hyper-mutator clone from MG1655 had two mutations in two different SOS genes, *uvrA* and *uvrC*, which are involved in DNA repair processes under stressed conditions [21] as well as an intragenic IS element mutation between *fnr* and *ogt*, where the later gene is a methyltransferase known to be involved in hypermutating phenotypes [22]. For the two DH1 hypermutator clones, it was not apparent which genes may have caused such phenotypes. Hypermutating clones were excluded from further analysis to simplify the genetic analysis and as they have a greater potential for instability when utilizing them as a platform strain.

Key mutations are presented in Table 2, which are defined as mutations in genes or regions that were found to be repeatedly mutated across different isolates of K-12 MG1655, DH1, or in both strains. A full summary of mutations for each isolate are given in Additional file 1. The mutations were categorized as 'combined' or 'strain-specific'. Overall, there were two genetic regions identified as 'combined' mutations occurring in both strains. Further, there were four and six strain-specific key mutations for MG1655 and DH1, respectively.

The first key mutation occurring in both strains was in the non-coding intergenic region between *mdtJ* and *tqsA*. Mutations in this region have been previously reported to improve tolerance of *E. coli* toward isobutanol [23]. The *mdtJ* gene encodes a component of a multidrug efflux pump that also physiologically exports spermidine [24]. Surprisingly, an identical deletion of Δ120 bp in the intergenic region, i.e., *mdtJI* promoter region, between the genes occurred in both of the strains. This deletion was found in every MG1655 clone isolated. The widespread penetration of this mutation could be due the fact that this deletion was the easiest to loop out under IL or other stress conditions [23] or could be due to occurrence of this mutation in the seeding culture for the experiment (although it did occur in both K-12 MG1655 and DH1). The other types of mutations in this region were structural changes in the *tqsA* gene—one was an intragenic in-frame Δ12 bp deletion, and the other was a Δ3035 bp deletion which included the *pntB*, and *pntA* genes located next to *tqsA* on the chromosome. The latter of these mutations is likely a loss of function mutation for *tqsA*. *PntA* and *pntB*, encode for the two subunits forming pyridine nucleotide transhydrogenase enzyme [25]

Table 2 Key mutations categorized by those which repeatedly mutated (i.e., had multiple unique mutations in any ORF or genetic region) in K-12 MG1655, DH1, or those which were shared across the strains

Strain	Gene	Mutation	Mutation type	Mutated allele function	Strain	IL observed	Count
Combined	mdtJ/tqsA	Intergenic (− 56/− 237) Δ120 bp	DEL	Transporter	MG	B, E	23
	tqsA/mdtJ	Intergenic (− 239/− 54) Δ120 bp	DEL	Transporter	DH1	B	5
	tqsA	Coding (857–868/1035 nt) Δ12 bp	DEL	Transporter	MG	B	6
	pntA–tqsA	Δ3035 bp	DEL	Transhydrogenase/transporter	DH1	B	2
	yhdP	Coding (2440–2443/3801 nt) Δ2 bp	DEL	Transporter	MG	B	1
		Coding (647/3801 nt) (TGGAGCC)1 → 2	INS	Transporter	MG	B	4
		Coding (200–201/3801 nt) IS5 (−) +4 bp	MOB	Transporter	MG	B	1
		Coding (3102–3110/3801 nt) IS element(+) +9 bp	MOB	Transporter	DH1	B	1
		Coding (2887–2890/3801 nt) Δ4 bp	DEL	Transporter	DH1	B	1
MG1655	rpoC	P359L (CCA → CTA)	SNP	RNA synthesis	MG	B	4
		F773Y (TTC → TAC)	SNP	RNA synthesis	MG	B	6
		R1075S (CGT → AGT)	SNP	RNA synthesis	MG	B	1
	cspC	Coding (33–41/210 nt) IS1 (−) +9 bp	MOB	Stress protein	MG	B	1
		Q37* (CAG → TAG)	SNP	Stress protein	MG	E	1
	rpsG	Coding (460/540 nt) Δ1 bp	DEL	Subunit of ribosome	MG	B	1
		L157* (TTA → TGA)	SNP	Subunit of ribosome	MG	B	2
	rph	Pseudogene (667/669 nt) + C	INS	Nucleotide biosynthesis	MG	E	2
	pyrE/rph	Δ82 bp	DEL	Nucleotide biosynthesis	MG	B	5
DH1	rho	G61E (GGA → GAA)	SNP	Transcription termination factor	DH1	B	2
		Y80H (TAC → CAC)	SNP	Transcription termination factor	DH1	E	5
		Y80C (TAC → TGC)	SNP	Transcription termination factor	DH1	B	2
		T406P (ACC → CCC)	SNP	Transcription termination factor	DH1	E	4
	fhuA	Coding (337–479/2244 nt) Δ143 bp	DEL	Transport of ferrichrome	DH1	E	1
		Coding (1442/2244 nt) (GTCATAACGACCGCCTAGGG)1 → 2	INS	Transport of ferrichrome	DH1	E	1
		Coding (2107/2244 nt) Δ1 bp	DEL	Transport of ferrichrome	DH1	B	2
		Coding (2129/2244 nt) Δ1 bp	DEL	Transport of ferrichrome	DH1	E	4
	rcdA	L55S (TTG → TCG)	SNP	Transcription regulator	DH1	E	1
		Coding (338–341/537 nt) IS element(+) +4 bp	MOB	Transcription regulator	DH1	E	1
	purB	K404T (AAG → ACG)	SNP	Nucleotide biosynthesis	DH1	B	2
		S21N (AGC → AAC)	SNP	Nucleotide biosynthesis	DH1	E	1
	gadE	Coding (273–281/528 nt) IS element(-) +9 bp	MOB	Transcriptional activator	DH1	E	2
		Coding (273–281/528 nt) IS element(+) +9 bp	MOB	Transcriptional activator	DH1	E	1

Different unique mutations in the same gene or allele were identified in different clones across the different experiments. The mutations were categorized as 'combined', i.e., identified in both strains, or 'strain-specific'. B and E denotes $[C_4C_1Im]Cl$ and $[C_2C_1Im][OAc]$ IL, respectively, where MG denotes K-12 MG1655 strain

and are important for redox balance in the cell [26, 27]. The *tqsA* gene encodes a transporter of quorum-sensing signal AI-2 which plays a role in control of biofilm formation in *E. coli* K-12 by enhancing transport of autoinducer-2 (AI-2) [28]. A Δ*tqsA* mutant was found to carry higher resistance to various drugs [28], which reveals a potentially tolerance role in the evolved strains in this study.

The second key mutation occurring in both strains was in *yhdP*, a gene encoding a putative transport protein [29]. A total of five unique mutations, all structural changes, were identified in *yhdP*—three in MG1655 and two in DH1. These mutations were two out-of-frame short deletions, two IS mobile element insertions, and a short 7 bp duplication. All of these structural mutations suggest a loss of function. There are no previous studies examining the role of *yhdP* in tolerance, to the best of our knowledge, making this finding a novel discovery.

Strain-specific key mutations (Table 2) were also identified in MG1655 and DH1. In MG1655, three different

coding mutations were identified in *rpoC*, encoding the β′ subunit of RNA polymerase. Prior ALE studies have identified *rpoC* coding mutations, which were found to both boost metabolic efficiency in glucose minimal medium [30] and improve growth at 42 °C [31, 32]. Probable loss-of-function mutations (premature stop codon and IS element insertion) were identified in *cspC* (encoding a stress protein of the CspA family). CspC is thought to stabilize *rpoS* mRNA when overexpressed [33] and to have activity as a transcription anti-terminator [34]. Mutations in this gene were previously found to play a role in stress responses [33]. Two different mutations occurred in the *rpsG* gene, encoding the essential S7 subunit of the 30S ribosome. These were a Δ1 bp deletion and a premature stop codon near the end of the gene. These mutations likely correct a defective 23 amino acid C-terminal extension to RpsG that occurs only in K-12 derived strains and that causes increased degradation of this protein. Similar mutations have previously been observed in MG1655 evolved for increased tolerance toward sodium cation [35]. Truncation of *rpsG* is thus likely a general stress coping mechanism. An Δ82 bp deletion was also found in the intragenic region between *pyrE* and *rph* and an insertion in *rph* was observed seven times in different clones. The *rph* gene encodes for an RNase PH [36], where *pyrE* encodes an orotate phosphoribosyltransferase [37]. Related deletion mutations were reported in different ALE studies including adaptation to lactate, minimal glucose medium, and high temperature (42 °C) [17, 30, 32]. The wild type strain *E. coli* K-12 has a frameshift mutation in *rph* which leads to pyrimidine starvation on minimal media due to resulting low levels of orotate phosphoribosyltransferase encoded by *pyrE* [38]. It appears that these mutations can be attributed to a 15% growth advantage by alleviation of defects in pyrimidines biosynthesis [39], and these mutations are predominantly general adaptations to growth on minimal medium. Interestingly, in DH1, mutations were not found in *rpoC*, *rpsG*, or *pyrE*/*rph*, with mutations in the latter two regions serving to correct metabolic and ribosomal protein defects that are present in all K-12 strains, including DH1.

Strain-specific mutations in DH1 included genes involved in the processes of transcriptional activation and transportation. In 13 isolates, *rho*, encoding the Rho transcription terminator with annotated function as transcription termination factor *Rho* [40], all contained coding SNPs. Coding mutations in Rho have previously been observed as a major contributor to ethanol tolerance [41, 42], and have been found to reduce the rate of Rho-dependent transcription termination in an ethanol-tolerant mutant [42]. The *fhuA* gene (encoding a ferrichrome outer membrane transporter) had several unique mutations in eight different isolates, including two unique Δ1 bp deletions, a Δ143 bp deletion, and a 20 bp short insertion. Additionally, two unique mobile element insertions were found between three isolates in *gadE* (encoding the GadE transcriptional activator), and a coding SNP and a mobile element insertion were found in two isolates in *rcdA* (encoding the RcdA transcriptional activator). Interestingly, deletion of *rcdA* was previously found to improve tolerance of DH1 toward IL [12]. Finally, two different coding SNPs in *purB* (encoding adenylosuccinate lyase) were found in three isolates.

Secondary screening of the evolved clones

A secondary screen was performed to generate quantitative data on the resequenced isolates after their genetic bases had been determined (see "Methods"). Resequenced isolates were clustered (see "Methods") based on their genotypes in order to assess their performance into three groups: genetically-identical where clones share identical genotypes; genetically-similar based on shared mutations (an expected outcome as multiple clones were isolated from the same population); and hyper-mutator isolates—which were eliminated from the secondary screening and further analysis. Overall, there were 3, 3, 2, and 3 genetically-similar clusters for the MG1655/$[C_4C_1Im]Cl$, DH1/$[C_4C_1Im]Cl$, MG1655/$[C_2C_1Im][OAc]$, and DH1/$[C_2C_1Im][OAc]$ conditions, respectively. Most of the clones showed reproducible performance when culturing. Results for both *E. coli* K-12 MG1655 and DH1 are summarized in Additional file 2: Tables S2 and S3, respectively. A few isolates did not grow during the secondary screen for unknown reasons (Additional file 2: Table S5).

Isolated strains from the study with similar genotypes exhibited a similar performance when tolerating ILs. A main difference in this study was that the growth rate criterion was used to quantify strains with improved performance. The coefficients of variation in growth rate (h^{-1}) between isolates that were genetically-identical or genetically-similar were 21 and 11%, respectively. Some of the resequenced isolates exhibited no growth (6 out of 45 clones, 4 with similar genotypes, Additional file 1). This non-growth could be a result of moving from unstressed to a highly-stressed condition during the screen, but it was not explored further. A more detailed analysis of the secondary screening results is provided in Additional file 2. The most promising isolates (based on criteria in Additional file 2: Figure S2) from each genetically identical or similar cluster were selected for further testing and are provided in Table 3.

Table 3 Selected clones from each genetically identical or similar cluster were selected for testing for tolerance to [C_2C_1Im][OAc] and [C_4C_1Im]Cl ionic liquids

TALEs	IL type	Concentration (%)	Gene set	Aver. growth rate (h^{-1})	Aver. final OD_{600}	Aver. lag-time (h)
MG 4.7	[C_4C_1Im]Cl	5.4	MG-BM-3A	0.26 ± 0.024	0.87 ± 0.039	4.88 ± 0.04
MG 3.10	[C_4C_1Im]Cl	5.4	MG-BM-3D	0.31 ± 0.009	0.94 ± 0.04	10.85 ± 1.24
MG 4.5	[C_4C_1Im]Cl	5.4	MG-BM-3E	0.30 ± 0.015	0.84 ± 0.082	3.19 ± 1.17
MG 11.10	[C_2C_1Im][OAc]	4.6	MG-EM-1	0.23 ± 0.003	1.69 ± 0.004	1.55 ± 0.006
MG 12.7	[C_2C_1Im][OAc]	4.6	MG-EM-1	0.27 ± 0.005	1.1 ± 0.03	1.9 ± 0.015
MG 10.9	[C_2C_1Im][OAc]	4.6	MG-EM-1A	0.09 ± 0.003	0.74 ± 0.004	1.99 ± 0.006
DH1 5.3	[C_4C_1Im]Cl	4.6	DH-BM-3A	0.32 ± 0.005	0.96 ± 0.01	12.51 ± 0.075
DH1 5.10	[C_4C_1Im]Cl	4.6	DH-BM-2	0.58 ± 0.023	0.61 ± 0.12	8.8 ± 1.059
DH1 6.7	[C_4C_1Im]Cl	4.6	DH-BM-1	0.36 ± 0.033	0.99 ± 0.094	5.62 ± 0.154
DH1 7.5	[C_4C_1Im]Cl	4.6	DH-BM-1	0.25 ± 0.025	1.39 ± 0.014	11.49 ± 3.604
DH1 13.10	[C_2C_1Im][OAc]	4.2	DH-EM-2A	0.14 ± 0.013	0.52 ± 0.385	9.59 ± 0.66
DH1 14.2	[C_2C_1Im][OAc]	4.2	DH-EM-2B	0.2 ± 0.027	0.36 ± 0.304	4.91 ± 1.083
DH1 13.8	[C_2C_1Im][OAc]	4.2	DH-EM-1	0.23 ± 0.007	0.6 ± 0.085	8.57 ± 3.393
DH1 15.2	[C_2C_1Im][OAc]	4.2	DH-EM-3C	0.28 ± 0.057	0.57 ± 0.148	6.33 ± 0.259
DH1 16.7	[C_2C_1Im][OAc]	4.2	DH-EM-3A	0.31 ± 0.014	0.26 ± 0.172	5.34 ± 0.44

Each of the TALE-derived isolates is presented with the corresponding IL-type and concentration in which it was originally evolved along with phenotypic characteristics of each of the selected isolates

Tolerance testing of the selected evolved strains and comparison to previous work

Fifteen isolates selected from the secondary screen results (Table 3) were tested for tolerance to [C_2C_1Im][OAc], an ionic liquid with arguably the best characteristics for lignocellulose solubilization and pretreatment [5]. This screen was performed to establish quantitative differences in final cellular densities achieved in batch culture (Fig. 2) and to understand cross-tolerance to other imidazolium-based ILs.

While almost all the isolates were capable of growth in M9 minimal medium in the presence of 4.3% (w/v) (250 mM) [C_2C_1Im][OAc] (Fig. 2a), a few isolates had significantly higher final densities in this condition. These high performing isolates were three K-12 MG1655 (MG 4.7, MG 3.10, and MG 4.5) and two DH1 (DH 5.10 and DH 15.2) derivatives. Interestingly, only one of the five best performers were actually evolved on [C_2C_1Im][OAc], DH1 15.2, while the remainder were isolated from the [C_4C_1Im]Cl evolutions. Furthermore, increased

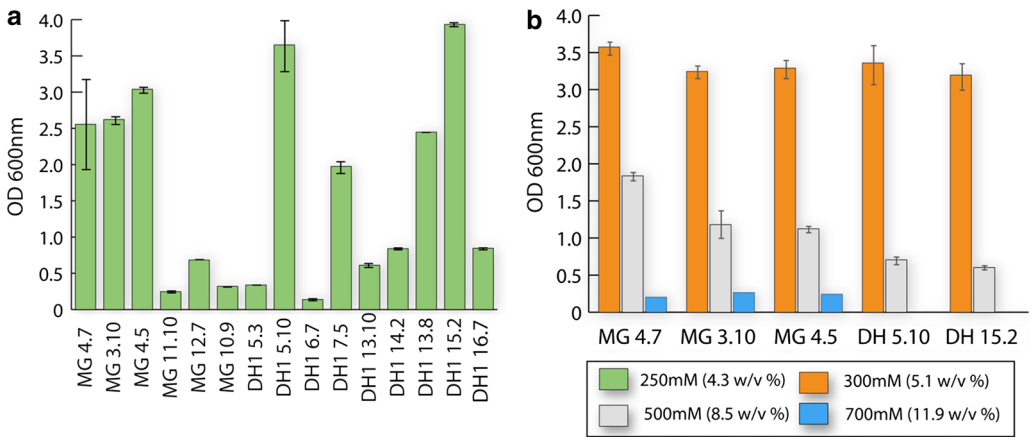


Fig. 2 Performance of evolved clones in batch culture under M9 minimal media conditions with various [C_2C_1Im][OAc] IL loadings. Shown is the final optical density (OD_{600nm}) of the selected best performing clones (Table 3) under different concentrations of [C_2C_1Im][OAc]: **a** first, a screen with 250 mM and all best performing clones, **b** second, a follow up screen with 300, 500 and 700 mM utilizing the highest final density clones from the 250 mM screen. Surprisingly, a number of clones that were originally evolved for [C_4C_1Im]Cl tolerance showed high cross-tolerance to elevated [C_2C_1Im][OAc] concentrations. At the highest concentration (700 mM), only the MG1655 derived strains showed measurable growth

amounts of $[C_2C_1Im][OAc]$ were tested with the five best performing isolates (Fig. 2b). Increasing amounts of IL inhibited growth in all the strains, but robust growth (i.e., a final $OD_{600} > 0.5$) was detected for all five clones at 8.5% (w/v) IL and detectable growth was observed for the MG clones, MG 4.7, MG 3.10, and MG 4.5, in concentrations up to approximately 11.9% (w/v) (700 mM) $[C_2C_1Im][OAc]$ (Fig. 2b). These tests demonstrated that the evolved isolates display cross tolerance to ILs which they were not exposed to during the ALE process and the levels of IL tolerated were impressive when compare to previously-developed strains.

The final concentration of ILs tolerated by evolved isolates using TALE were compared to previously reported strains generated for IL tolerance (Table 4). The robust growth of the best performing clones MG 4.7, MG 3.10, DH 5.10 and DH 15.2 observed at 8.5% (w/v) compares favorably with other reported values for engineered and evolved strains. For example, the tolerance achieved from [12] was based on introducing a mutation in the transcriptional regulator encoded by *rcdA* with tolerance up to 3% (w/v) $[C_2C_1Im][OAc]$ achieved in LB medium. Additionally, thermophilic communities have been isolated by enriching them for tolerance to $[C_2C_1Im][OAc]$, which has resulted in the identification of a mixed population tolerant to 6% w/v $[C_2C_1Im][OAc]$ [10]. Finally, an ALE approach had also been previously employed to develop a strain tolerant to $[C_4C_1Im]Cl$ of approximately 7% (w/v) in rich media (LB) [11]. A direct experimental comparison to two previously developed strains was also conducted.

Comparison of selected evolved strains to previously-developed tolerant strains

The best performing IL tolerant isolates, MG 4.7 and MG 3.10, were compared to rationally-engineered IL tolerant strains, JBEI-10101 [8] and JBEI-13314 [12], to provide a direct comparison for the efficacy of the evolution process as compared to rational engineering approaches. JBEI-10101 [8] is DH1 harboring a plasmid containing genes for an MFS-1 pump from *Enterobacter lignolyticus* and its response regulator (*eilAR*), and the JBEI-13314 [12] is DH1 carrying a deletion in *rcdA*, which encodes a predicted transcriptional regulator of the MFS-1 pump *ybjJ*. Two different media types were used in this comparison: a rich undefined LB medium and a minimal defined M9-glucose medium. This comparison was performed using two different promising IL compounds, 300 mM of either $[C_2C_1Im][OAc]$ [5.1% (w/v)] or $[C_2C_1Im][Cl]$ [4.4% (w/v)]. $[C_2C_1Im][Cl]$ which contains a chloride anion, was found to be effective towards dissolving cellulose in comparison to larger anions, $[C_2C_1Im][OAc]$ [43]. Furthermore, ILs with anions such as acetate (e.g., $[C_2C_1Im][OAc]$) have lower viscosities and this is beneficial as it facilitates the dissolution process [9].

The performance of the TALE derived strains was superior to those developed through rational engineering. In LB medium at 300 mM of $[C_2C_1Im][OAc]$ or $[C_2C_1Im]Cl$, the performance of JBEI-13314 and JBEI-10101 was improved over the background control of a wild-type MG1655. In the same conditions, the TALE-derived strains, MG 4.7 and MG 3.10, grew at a significantly faster rate and to a higher final density than the

Table 4 Comparison of IL tolerance in the generated TALE evolved strains in the current study and previously reported tolerances from different studies with each respective tolerant biological system

Strain	IL's type	Description/mechanism	Culturing environment	Concentration % (w/v) criteria	References
E. coli K-12 MG1655 ΔfadD mutant	$[C_4C_1Im]Cl$	ALE (90 days long)	LB broth	7.0	[11]
E. coli DH1	$[C_2C_1Im]Cl$	Efflux pump encoded by *eilA* from *Enterobacter lignolyticus*	Minimal media (M9)	5.8	[9]
JBEI-10101 (*E. coli DH1*)	$[C_2C_1Im]Cl$	Native *eilA* pump from *Enterobacter lignolyticus*	Minimal media (M9)	5.5	[8]
Thermophilic communities	$[C_2C_1Im][OAc]$	*Thermophilic* enrichment	Raw material with minimal media	6.0	[10]
JBEI-13314 (*E. coli* DH1 rcdA mutant)	$[C_2C_1Im][OAc]$	Mutation of transcriptional regulator encoded by rcdA	LB broth	3.0	[12]
E. coli K-12 MG1655 mutants MG4.7, MG 3.10, and MG 4.5	$[C_2C_1Im][OAc]$	TALE isolate	Minimal media (M9)	8.5	This work
E. coli DH1 mutants DH 5.10 and 15.2	$[C_2C_1Im][OAc]$	TALE isolate	Minimal media (M9)	8.5	This work

The high performing isolates from this work were three K-12 MG1655 mutants (MG 4.7, MG 3.10, and MG 4.5) and two DH1 mutants (DH 5.10 and DH 15.2)

rationally engineered strains. Specifically, the TALE derived strain, MG 4.7, grew to a final OD_{600} of approximately 3.0 and 1.5 in LB containing $[C_2C_1Im][OAc]$ and $[C_2C_1Im]Cl$, respectively, versus final ODs of approximately 1.0 on both ILs for the best performing rationally engineered JBEI-10101 strain. Similarly, the growth rates were higher for the TALE-derived strains (Fig. 3a, b). It should be noted that the TALE-derived strains had not been evolved in LB, whereas the JBEI strains were benchmarked in LB (or similar rich medium) as a base medium [8, 12]. Further, in M9 glucose medium containing 300 mM $[C_2C_1Im][OAc]$ or 300 mM $[C_2C_1Im]Cl$, the final cell biomass, OD_{600}, values reached were approximately, 1.5 and 4.0 for MG 4.7 and 1.0 and 3.0 for MG 3.10 in $[C_2C_1Im][OAc]$ and $[C_2C_1Im]Cl$, respectively (Fig. 3c, d), whereas JBEI-13314 and JBEI-10101 exhibited virtually no detectable growth.

The tolerance phenotypes demonstrated by the TALE-derived isolates gives a strong indication of the causality of the mutations identified in this work, specifically, the 'combined' key mutations (Table 2). Both the MG 4.7 and MG 3.10 were found to carry ten mutations each, eight being identical and shared, and several mutations were from the combined key mutation set. Namely, MG 4.7 and MG 3.10 share a Δ120 bp deletion in the intergenic region between *mdtJ*/*tqsA* and 12 bp deletion in *tqsA*. Both strains also carry an additional key mutation, a coding SNP in *rpoC*. Further, they both carry different frameshift mutations in the *yhdP* gene. Differentiating key mutations are a Δ1 bp deletion in *rpsG* in MG 3.10 isolate, and a coding SNP in *rpsA* in MG 4.7, with both genes encoding ribosomal protein subunits.

The best performing DH1 isolates, DH1 5.10 and DH1 15.2, also shared key mutations. Each isolate possessed seven mutations overall, none of them identical between the isolates. However, shared genes that were mutated included coding SNPs in the *rho* gene. Other strain-specific key mutations were identified in the strains including SNPs in *purB* and *cspC* identified in DH1 5.10 and a frameshift insertion mutation in *fhuA* in DH1 15.2. It should be noted that the gene encoding the regulator of *purB*, *purR* [44], carried a mutation in the isolate that did

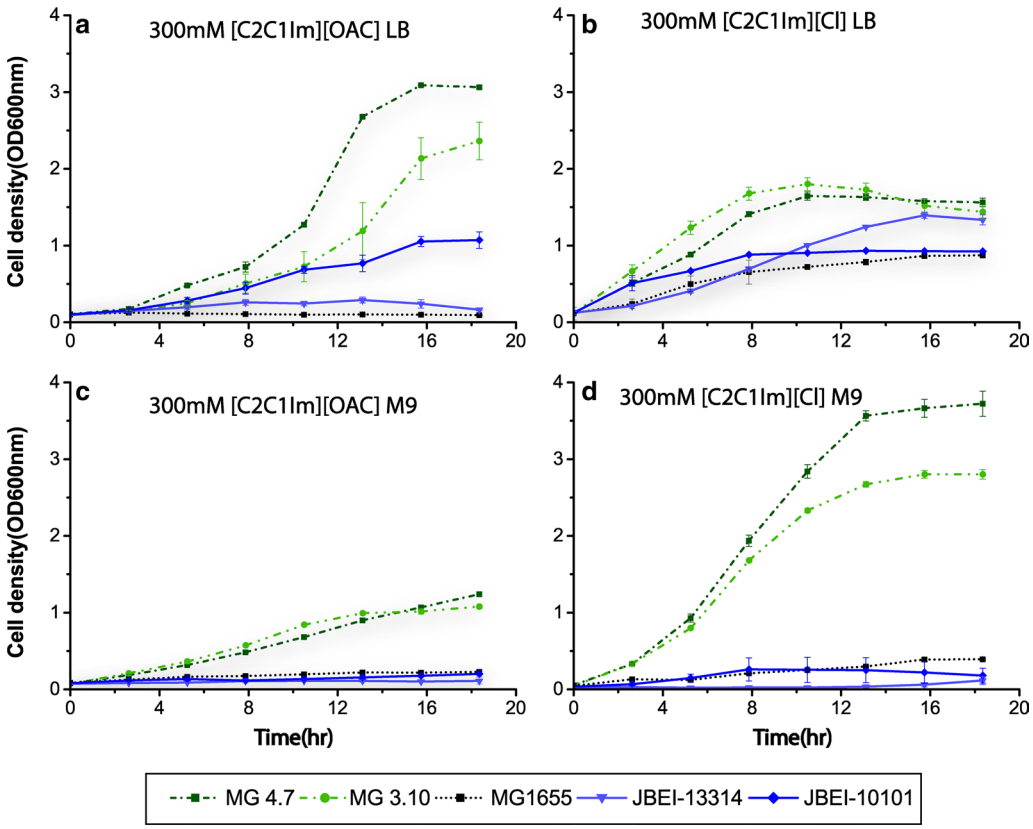


Fig. 3 Comparison of TALE evolved IL tolerant clones MG1655#4.7 and MG1655#3.10 to previously engineering tolerant strains JBEI-13314 and JBEI-10101 in LB (**a**, **b**) and M9 (**c**, **d**) media containing either 300 mM of $[C_2C_1Im][OAc]$ [5.1% (w/v)] or $[C_2C_1Im][Cl]$ [4.4% (w/v)]. TALE evolved clones exhibited improved growth compared to rationally-designed strains, particularly in M9, where JBEI-13314 and JBEI-10101 were severely inhibited

not possess a direct mutation in *purB*. Thus, the mutations identified as key in Table 2 can be linked to high performing phenotypes and are likely causal. However, detailed studies to reveal their mechanism of causality are required.

Discussion

The economical and efficient break down of lignocellulosic material into carbon feedstocks is an essential step in renewable bioprocessing. Ionic liquid (IL) solubilization is a promising method for breakdown of lignocellulosic material, however these compounds are toxic to most bioproduction chassis strains. Thus, the scope of this study was to generate IL tolerant strains utilizing an adaptive laboratory evolution process. Accordingly, the main contributions from this work are: (1) effective generation of IL tolerant strains (including cross-tolerance) for two common production chassis, *E. coli* K-12 MG1655 and DH1, which can be used as platform strains for utilizing feedstocks generated through IL degradation methods, (2) insights into both strain-specific and global mechanisms of IL tolerance through examining key mutations found in multiple parallel evolved isolates, and (3) establishing a viable method using a multiple population TALE approach with next-generation sequencing towards generating tolerant strains. This method was benchmarked via comparison to rational engineering approaches.

TALE was successful in generating strains that were tolerant to the targeted ILs. After approximately 40 days of continuous exposure to ILs during growth (mostly in exponential growth), populations of cells were able to grow at approximately threefold or greater of the initial concentration of each IL compared to the wild type (Table 1). Tolerance levels of isolated clones are impressive when compared to other tolerant bacterial strains [8, 12, 45]. Additionally, the selected best performing strains demonstrated high level of IL cross-tolerance toward $[C_2C_1Im][OAc]$; detectable growth at 11.9% (w/v) for TALE-derived *E. coli* K-12 MG1655 clones and robust growth at 8.5% (w/v) for the same MG1655 isolates plus the *E. coli* DH1 best-performing clones. Thus, the TALE-derived strains show promise as platform strains for utilization of biomass hydrolysates generated using IL treatment.

The key mutations identified from this study provide insights into the potential mechanisms of tolerance phenotypes in the evolved strains. The most prevalent and shared mutations observed were in the *mdtJ*/*tqsA* intergenic region, as well as in the *tqsA* gene, and in the *yhdP* gene (Fig. 4). The key mutations identified in this study were specific to ILs when compared to a control experiment where K-12 MG1655 was evolved on M9 glucose minimal medium at the same temperature but without any stress from ILs [17]. Thus, it appears that modulating transport, likely of ILs, in and out of the cell is crucial for tolerance to the ILs tested here and likely similar compounds. This finding further supports the focus of rationally modulating transport systems in engineering tolerance [12, 23, 46]. However, identifying which transporters are critical for tolerance in a given strain de novo is difficult, therefore making the use of TALE a powerful approach.

A genetic-level analysis of the specific multidrug transport system mutation observed in both strains provides a glimpse into the mechanistic impact of these mutations. The most prevalent Δ120 bp deletion in the intergenic region between *mdtJ* and *tqsA* likely disrupts H-NS binding sites in the *mdtJI* promotor region, which is then believed to relieve negative repression of transcription of *mdtJI* by H-NS, given its role as a steric hindrance protein [47]. Supporting this, when H-NS is absent, a ninefold increase in *mdtJI* expression occurs as compared to wild type *E. coli*, and consequently the activity of MdtJI as well [47]. This finding further highlights the likely active role of this small multi-drug resistance (SMR) efflux pump in the resistance mechanism. Similar work has shown such pumps to be active on a wide range of inhibitory compounds [23, 46–48]. It is noteworthy that in the control experiment [17], an intergenic *hns*/*tdk* mutation was reported in almost all evolved endpoints where *hns* was determined to be upregulated and conferred a fitness advantage, i.e., fast growth rate, likely through subsequent downregulation of stress responses. Given that no similar intergenic *hns*/*tdk* mutations were seen in this work with ILs present during the evolution, this finding further supports the importance of high expression of *mdtJI* towards tolerance of the ILs examined here and the benefit of control ALE experiments.

The other key mutated region, in this case one gene, identified in the tolerant clones was in *yhdP*. The *yhdP* gene encodes a predicted transporter [29]. The occurrence of five unique mutations, all interpreted to be loss of function mutations, imply that removing this gene is a viable strategy for increased IL tolerance. However, the specific mechanism is unclear as to what metabolite is pumped in or out of the cell to provide the increased fitness. One can speculate that ILs could enter through this transporter, but this has yet to be verified. Future work could include an effort to definitively assign the causality of key mutations. For example, expression profiling could be performed on isolates or reconstructed strains carrying only the Δ120 bp deletion in the intergenic region between *mdtJ* and *tqsA*. Such transcription levels could help focus on the impact of *mdtJ* and/or *tqsA* and lead to

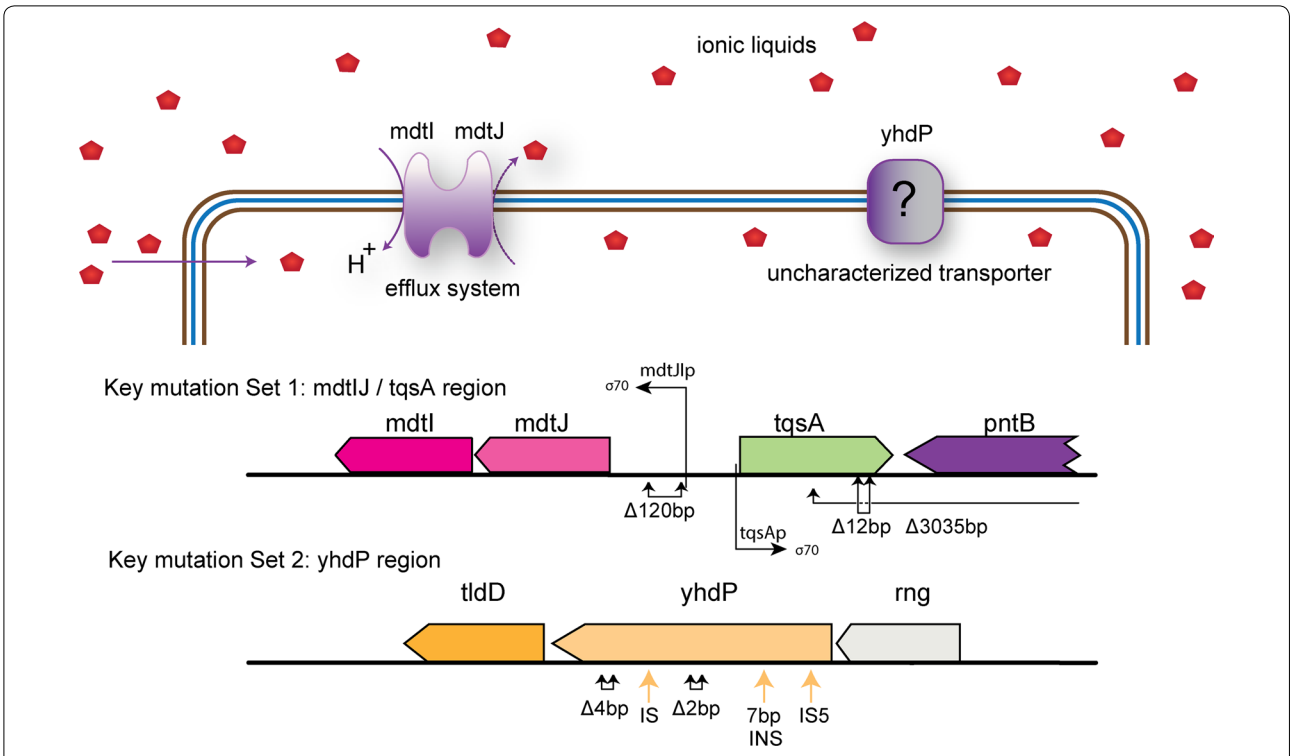


Fig. 4 A diagram of the cell showing processes associated with key mutations. A cartoon diagram of key mutations in potential causal genetic regions identified in the evolved strains. A total of eight mutations are represented amongst different genetic regions. A Δ120 bp deletion was found in the non-coding region of *mdtJ*, near the promoter, and its neighboring gene *tqsA*. Two other structural changes were found in the *tqsA* gene—one was an intragenic in-frame Δ12 bp deletion, the other was a Δ3035 bp deletion which included a major section of *tqsA* and the *pntB* and *pntA* genes located next to *tqsA* on the chromosome. Finally, five structural changes were identified in *yhdP* gene. These were two out-of-frame short deletions (Δ2 bp and Δ4 bp), two intergenic IS mobile element insertions, and a short 7 bp insertion/duplication. These mutations indicate a probable loss-of-function of *yhdP*

a better understanding of the underlying mechanisms of tolerance.

The TALE approach of independently passaging multiple populations in an automated, strictly-controlled platform, coupled with next-generation sequencing resulted in an effective process for generating tolerant strains and for revealing the key causal mutations. Sequencing the whole genome revealed mutational changes in the evolved strains when compared to the reference strains. However, relating a specific mutation or a set of mutations, i.e., genotype, to the apparent phenotype in certain conditions is time-consuming [39, 49, 50]. The use of multiple independent replicates allowed for the identification of mutations in the same gene or genetic region multiple times across different TALE experiments. This approach of using many replicates to decipher the causality of a mutation or set of mutations in a given strain appeared effective given that the best performing clones, MG4.7 and MG3.10, possessed such key shared mutations. To validate the key mutations identified, as well as to confirm the efficacy of the evolution process, the selected clones were compared to rationally-designed strains in two different media with closely similar ILs. The performance of TALE-derived strains was superior, which indicated the efficacy of utilizing TALE and pointed to the identified mutations in the generated the strains.

In summary, utilizing the TALE approach outlined here to generate IL-tolerant strains resulted in the generation of promising platform strains with enhanced tolerance toward high concentrations of ILs [up to 11.9% (w/v)]. The approach used to identify and interpret the key causal mutations using whole genome sequencing complemented with analyzing isolates from multiple independent populations, and multiple isolates from each population, was successful in revealing the key mutations involved in IL tolerance phenotypes. The most striking identified key mutations appeared to involve modulation of transport mechanisms, possibly the direct transport of ILs into and out of the cell. The results of this study and the approach used to generate tolerant strains can be expanded to other conditions, strains, and selection criteria, which would help in fast-tracking the utilization of alternative renewable feedstocks, as well as to

better understand tolerance mechanisms for inhibitory compounds.

Additional files

Additional file 1. All of the whole genome sequencing results.

Additional file 2. Additional text, tables and figures.

Abbreviations

IL: ionic liquid; TALE: tolerance adaptive laboratory evolution; ALE: adaptive laboratory evolution; *E. coli*: *Escherichia coli*; SMR: small multi-drug resistance.

Authors' contributions

Conceived and designed the experiments: ETTM and AMF. Performed most of the experiments: ETM. Performed comparison with rationally-designed strains and cross-tolerance: SW. Analyzed the data: ETM, AMF, RML, MJH and SWS. Contributed reagents/materials/analysis tools: BAS. Wrote the paper: ETM and AMF. RML, MJH, BAS and SWS reviewed manuscript and helped in discussing the results. All authors read and approved the final manuscript.

Author details

[1] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Building 220, Kemitorvet, 2800 Kgs. Lyngby, Denmark. [2] Department of Bioengineering, University of California, 9500 Gilman Drive La Jolla, San Diego, CA 92093, USA. [3] Joint Bioenergy Institute, Emeryville, CA, USA. [4] Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [5] State Key Laboratory of Chemical Resource Engineering, College of Life Science and Technology, Beijing University of Chemical Technology, Beijing 100029, People's Republic of China.

Acknowledgements

We would like to thank Mohammad Radi, Anna Koza, Patrick Phaneuf, Professor Gang Cheng, and Professor Qipeng Yuan for their help in the project.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All data generated or analyzed during this study are included in this published article (and its Additional files).

Consent for publication

This article does not contain any studies with animals or with human participants performed by any of the authors.

Ethics approval and consent to participate

Not applicable.

Funding

Funding for the project was provided by The Novo Nordisk Foundation Grant Numbers NNF14OC0011269 and NNF10CC1016517. Portions of this work were supported by International Joint Graduate-Training Program of Beijing University of Chemical Technology. Part of this work was part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

Publisher's Note

Received: 25 August 2017 Accepted: 9 November 2017
Published online: 16 November 2017

References

1. Johnson JM, Coleman MD, Gesch R, Jaradat A, Mitchell R, Reicosky D, Wilhelm WW. Biomass, bioenergy crops in the United States: a changing paradigm. Am J Plant Sci Biotechnol. 2007;1(1):1–28.
2. Hoffert MI, Caldeira K, Benford G, Criswell DR, Green C, Herzog H, et al. Advanced technology paths to global climate stability: energy for a greenhouse planet. Science. 2002;298:981–7.
3. Moe ST, Janga KK, Hertzberg T, Hägg MB, Øyaas K, Dyrset N. Saccharification of lignocellulosic biomass for biofuel and biorefinery applications A renaissance for the concentrated acid hydrolysis? Energy Procedia. 2012;20:50–8. https://doi.org/10.1016/j.egypro.2012.03.007.
4. Jönsson LJ, Martín C. Pretreatment of lignocellulose: formation of inhibitory by-products and strategies for minimizing their effects. Bioresour Technol. 2016;199:103–12.
5. Sun N, Liu H, Sathitsuksanoh N, Stavila V, Sawant M, Bonito A, et al. Production and extraction of sugars from switchgrass hydrolyzed in ionic liquids. Biotechnol Biofuels. 2013;6:39. http://biotechnologyforbiofuels.biomedcentral.com/articles/10.1186/1754-6834-6-39. Accessed 13 Oct 2017.
6. Li C, Tanjore D, He W, Wong J, Gardner JL, Sale KL, et al. Scale-up and evaluation of high solid ionic liquid pretreatment and enzymatic hydrolysis of switchgrass. Biotechnol Biofuels. 2013;6:154.
7. Li C, Knierim B, Manisseri C, Arora R, Scheller HV, Auer M, et al. Comparison of dilute acid and ionic liquid pretreatment of switchgrass: biomass recalcitrance, delignification and enzymatic saccharification. Bioresour Technol. 2010;101:4900–6. https://doi.org/10.1016/j.biortech.2009.10.066.
8. Ruegg TL, Kim E-M, Simmons BA, Keasling JD, Singer SW, Lee TS, et al. An auto-inducible mechanism for ionic liquid resistance in microbial biofuel production. Nat Commun. 2014;5:3490.
9. Frederix M, Hütter K, Leu J, Batth TS, Turner WJ, Rüegg TL, et al. Development of a native *Escherichia coli* induction system for ionic liquid tolerance. PLoS ONE. 2014;9:1–10.
10. Reddy AP, Simmons CW, Claypool J, Jabusch L, Burd H, Hadi MZ, et al. Thermophilic enrichment of microbial communities in the presence of the ionic liquid 1-ethyl-3-methylimidazolium acetate. J Appl Microbiol. 2012;113:1362–70.
11. Qin D, Hu Y, Cheng J, Wang N, Li S, Wang D. An auto-inducible *Escherichia coli* strain obtained by adaptive laboratory evolution for fatty acid synthesis from ionic liquid-treated bamboo hydrolysate. Bioresour Technol. 2016;221:375–84.
12. Frederix M, Mingardon F, Hu M, Sun N, Pray T, Singh S, et al. Development of an *E. coli* strain for one-pot biofuel production from ionic liquid pretreated cellulose and switchgrass. Green Chem. 2016;18:4189–97.
13. LaCroix RA, Palsson BO, Feist AM. A model for designing adaptive laboratory evolution experiments. Appl Environ Microbiol. 2017;83:AEM.03115–6. https://doi.org/10.1128/AEM.03115-16.
14. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol. 2014;1151:165–88.
15. Baeshen MN, Al-Hejin AM, Bora RS, Ahmed MMM, Ramadan HAI, Saini KS, et al. Production of biopharmaceuticals in *E. Coli*: current scenario and future perspectives. J Microbiol Biotechnol. 2015;25:953–62.
16. Monk JM, Koza A, Campodonico MA, Machado D, Seoane JM, Palsson BO, et al. Multi-omics quantification of species variation of *Escherichia coli* links molecular features with strain phenotypes. Cell Syst. 2016;3(238–251):e12.
17. LaCroix RA, Sandberg TE, O'Brien EJ, Utrilla J, Ebrahim A, Guzman GI, et al. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of *Escherichia coli* K-12 MG1655 on glucose minimal medium. Appl Environ Microbiol. 2015;81:17–30. https://doi.org/10.1128/AEM.02246-14.
18. Lan W, Liu CF, Sun RC. Fractionation of bagasse into cellulose, hemicelluloses, and lignin with ionic liquid treatment followed by alkaline extraction. J Agric Food Chem. 2011;59:8691–701. https://doi.org/10.1021/jf201508g.

19. Lee SH, Doherty TV, Linhardt RJ, Dordick JS. Ionic liquid-mediated selective extraction of lignin from wood leading to enhanced enzymatic cellulose hydrolysis. Biotechnol Bioeng. 2009;102:1368–76. https://doi.org/10.1002/bit.22179.
20. Lee DH, Feist AM, Barrett CL, Palsson B. Cumulative number of cell divisions as a meaningful timescale for adaptive laboratory evolution of *Escherichia coli*. PLoS ONE. 2011;6:e26172. https://doi.org/10.1371/journal.pone.0026172.
21. Lin JJ, Sancar A. A new mechanism for repairing oxidative damage to DNA: (A)BC excinuclease removes AP sites and thymine glycols from DNA. Biochemistry. 1989;28:7979–84.
22. Horst JP, Wu TH, Marinus MG. *Escherichia coli* mutator genes. Trends Microbiol. 1999;7:29–36.
23. Minty JJ, Lesnefsky AA, Lin F, Chen Y, Zaroff TA, Veloso AB, et al. Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in *Escherichia coli*. Microb Cell Fact. 2011;10:18. http://microbialcellfactories.biomedcentral.com/articles/10.1186/1475-2859-10-18. Accessed 22 Sept 2016.
24. Higashi K, Ishigure H, Demizu R, Uemura T, Nishino K, Yamaguchi A, et al. Identification of a spermidine excretion protein complex (MdtJI) in *Escherichia coli*. J Bacteriol. 2008;190:872–8.
25. Clarke DM, Loo TW, Gillam S, Bragg PD. Nucleotide sequence of the pntA and pntB genes encoding the pyridine nucleotide transhydrogenase of *Escherichia coli*. Eur J Biochem. 1986;158:647–53.
26. Haverkorn van Rijsewijk BRB, Kochanowski K, Heinemann M, Sauer U. Distinct transcriptional regulation of the two *Escherichia coli* transhydrogenases PntAB and UdhA. Microbiology. 2016;162:1672–9.
27. Sauer U, Canonaco F, Heri S, Perrenoud A, Fischer E. The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. J Biol Chem. 2004;279:6613–9.
28. Herzberg M, Kaye IK, Peti W, Wood TK. YdgG (TqsA) controls biofilm formation in *Escherichia coli* K-12 through autoinducer 2 transport. J Bacteriol. 2006;188:587–98.
29. Köstner M, Schmidt B, Bertram R, Hillen W. Generating tetracycline-inducible auxotrophy in *Escherichia coli* and *Salmonella enterica* serovar typhimurium by using an insertion element and a hyperactive transposase. Appl Environ Microbiol. 2006;72:4717–25.
30. Conrad TM, Frazier M, Joyce AR, Cho B-K, Knight EM, Lewis NE, et al. RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. Proc Natl Acad Sci USA. 2010;107:20500–5.
31. Nedea EC, Markov D, Naryshkina T, Severinov K. Localization of *Escherichia coli* rpoC mutations that affect RNA polymerase assembly and activity at high temperature. J Bacteriol. 1999;181:2663–5.
32. Sandberg TE, Pedersen M, Lacroix RA, Ebrahim A, Bonde M, Herrgard MJ, et al. Evolution of *Escherichia coli* to 42 °C and subsequent genetic engineering reveals adaptive mechanisms and novel mutations. Mol Biol Evol. 2014;31:1–38. https://doi.org/10.1093/molbev/msu209.
33. Phadtare S, Inouye M. Role of CspC and CspE in regulation of expression of RpoS and UspA, the stress response proteins in *Escherichia coli*. J Bacteriol. 2001;183:1205–14.
34. Bae W, Xia B, Inouye M, Severinov K. *Escherichia coli* CspA-family RNA chaperones are transcription antiterminators. Proc Natl Acad Sci USA. 2000;97:7784–9.
35. Wu X, Altman R, Eiteman MA, Altman E. Adaptation of *Escherichia coli* to elevated sodium concentrations increases cation tolerance and enables greater lactic acid production. Appl Environ Microbiol. 2014;80:2880–8.
36. Kelly KO, Deutscher MP. Characterization of *Escherichia coli* RNase PH. J Biol Chem. 1992;267:17153–8.
37. Poulsen P, Bonekamp F, Jensen KF. Structure of the *Escherichia coli* pyrE operon and control of pyrE expression by a UTP modulated intercistronic attentuation. Embo J. 1984;3:1783–90.
38. Jensen KF. The *Escherichia coli* K-12 "wild types" W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. J Bacteriol. 1993;175:3401–7.
39. Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, et al. Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol. 2009;10:R118. http://scholarscompass.vcu.edu/cmsc_pubs. Accessed 17 Oct 2016.
40. Roberts JW. Termination factor for RNA synthesis. Nature. 1969;224:1168–74.
41. Goodarzi H, Bennett BD, Amini S, Reaves ML, Hottes AK, Rabinowitz JD, et al. Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*. Mol Syst Biol. 2010;6:378.
42. Haft RJF, Keating DH, Schwaegler T, Schwalbach MS, Vinokur J, Tremaine M, et al. Correcting direct effects of ethanol on translation and transcription machinery confers ethanol tolerance in bacteria. Proc Natl Acad Sci. 2014;111:E2576–85. https://doi.org/10.1073/pnas.1401853111.
43. Isik M, Sardon H, Mecerreyes D. Ionic liquids and cellulose: dissolution, chemical modification and preparation of new cellulosic materials. Int J Mol Sci. 2014;15:11922–40.
44. He B, Smith JM, Zalkin H. *Escherichia coli* purB gene: cloning, nucleotide sequence, and regulation by purR. J Bacteriol. 1992;174:130–6.
45. Yu C, Simmons BA, Singer SW, Thelen MP, VanderGheynst JS. Ionic liquid-tolerant microorganisms and microbial communities for lignocellulose conversion to bioproducts. Appl Microbiol Biotechnol. 2016;100:10237–49. https://doi.org/10.1007/s00253-016-7955-0.
46. Bay DC, Stremick CA, Slipski CJ, Turner RJ. Secondary multidrug efflux pump mutants alter *Escherichia coli* biofilm growth in the presence of cationic antimicrobial compounds. Res Microbiol. 2017;168:208–21. https://doi.org/10.1016/j.resmic.2016.11.003.
47. Leuzzi A, Di Martino ML, Campilongo R, Falconi M, Barbagallo M, Marcocci L, et al. Multifactor regulation of the MdtJI polyamine transporter in *Shigella*. PLoS ONE. 2015;10:1–19.
48. Yung PY, Lo Grasso L, Mohidin AF, Acerbi E, Hinks J, Seviour T, et al. Global transcriptomic responses of *Escherichia coli* K-12 to volatile organic compounds. Sci Rep. 2016;6:19899.
49. Herring CD, Glasner JD, Blattner FR. Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. Gene. 2003;311:153–63.
50. Utrilla J, O'Brien EJ, Chen K, McCloskey D, Cheung J, Wang H, et al. Global rebalancing of cellular resources by pleiotropic point mutations illustrates a multi-scale mechanism of adaptive evolution. Cell Syst. 2016;2:260–71. https://doi.org/10.1016/j.cels.2016.04.003.

Paper 3: Generation of bacterial platform for aromatic acid tolerance and utilization using adaptive laboratory evolution (under preparation)

Title

Generation of bacterial platform strains for aromatic acid tolerance using adaptive laboratory evolution

Authors

Elsayed T. Mohamed[1], Davinia Salvachúa[2], Christine Singer[2], Kiki Szostkiewicz, Manuel Jimenez-Diaz[3,5], Thomas Eng[3,5], Mohammad S. Radi[1], Aindrila Mukhopadhyay[3,5], Markus J. Herrgård[1], Steven W. Singer[3,5], Gregg T. Beckham[2], Adam M. Feist[1,4]

Affiliations

[1]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Building 220, Kemitorvet, 2800 Kgs. Lyngby, Denmark.

[2]National Renewable Energy Laboratory, Lakewood, CO 80401, USA

[3]Joint Bioenergy Institute, Emeryville, CA, USA.

[4]Department of Bioengineering, University of California, 9500 Gilman, Drive La Jolla, San Diego, CA 92093, USA

[5]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Abstract

Pseudomonas putida is a promising production chassis, as it has the capacity to metabolize aromatic compounds, including those derived from lignocellulosic biomass. However, the lignocellulosic biomass degradation process typically yields high titers of aromatic acids at levels that are toxic for *Pseudomonas putida*. In order to generate strains capable of utilizing and tolerating high concentrations of aromatic compounds, adaptive laboratory evolution was utilized to improve growth properties of *Pseudomonas putida* KT2440 (KT2440) strain in the presence of two common lignocellulosic feedstock breakdown aromatic acids, *p*-coumaric and ferulic acid. Additionally, KT2440 was evolved for higher fitness on glucose as a control experiment. For multiple independent replicates of KT2440 strain, cells were repeatedly passed to select for improved tolerance as well for fitness over the course of approximately 500 generations for aromatic acids tolerance and 912 generations for the control experiment on glucose using an automated liquid handler platform. The best performing strains were capable of growing on aromatic acids at a growth rate of (0.37 hr^{-1} compared to 0.17 hr^{-1}), (0.35 hr^{-1} compared to 0.16 hr-1), (0.68 hr^{-1} compared to 0.36 hr-1) for the wild type strain on 20 g/L *p*-coumarate, 30 g/L ferulate and 10 g/L of the mixture (*p*-coumarate, ferulate 1:1), respectively. Fitness improved strains on glucose were able to grow at a growth rate up to 0.84 hr^{-1} compared to 0.53 hr^{-1} for the wild type KT2440. Intermediate and end-point clonal and population samples were subjected to whole genome sequencing in order to reveal reproducibly occurring mutations and link them to observed phenotypes. The most reproducibly mutated gene were in *ttgB*, *oprB*-II, *algE*, rho, *gacS*, *accA*, *accD, sucA,* uvrY and the intergenic region between PP_5245/*kefB-III*. Most of the identified genes were involved in membrane transport mechanisms, central carbon metabolism, cell motility, and global transcriptional and post-transcriptional regulation processes. The causality of the mutations related to the gene products of hypothetical protein predicted to be the outer membrane protein, *algE* and a subunit of a membrane efflux pump, *ttgB* were confirmed via gene deletion experiments. It was found that *algE* deletion confers a growth benefit compared to the starting strain of (1.37-1.75)-fold increase in terms of growth rate with considerably reduced lag times. Contrary, *ttgB* gene deletion showed a negative impact on the growth of adapted KT2440 strains which underlines its importance for tolerance on aromatic acids. Overall, this work demonstrated that ALE can generate platform host strains for effectively metabolizing concentrated

lignocellulosic biomass breakdown monomers and identified causal mutations enabling such phenotypes present an engineering part to develop tolerance in other *P. putida* production strains.

Keywords

Pseudomonas putida KT2440, adaptive laboratory evolution, lignocellulosic feedstock, *p*-coumaric acid, ferulic acid.

Introduction

The clear need for an economically feasible industrial bioproduction-based on microbial cell factories can be met through the utilization of renewable feedstocks as a carbon and energy sources. Utilization of the feedstocks can facilitate building a sustainable economy by replacing part of the energy and chemical produced from fossil fuels. Lignocellulosic biomass possesses a worthy opportunity as the feedstock. Although most biomass has intrinsic low energy density [1], this issue can be compensated by large quantities of lignocellulosic biomass that are available at low costs from agricultural and forestry waste [2]. The facts give a strong motivation to develop economically feasible conversion processes of lignocellulosic biomass into chemicals, biomaterials, energy.

Among carbon components obtainable from lignocellulosic biomass, lignin remains an untapped resource for chemical and fuel production unlike other sugars (e.g., glucose, xylose, galactose, etc.) [3] because its structural heterogeneity and recalcitrance have limited its usage [4,5]. However, lignin typically presents from 15% to around 40% of the plant biomass which is a significant amount to be used; up to 40% of major monomeric building blocks in lignin [6] are *p*-coumaric acid (*p*-CA) and *trans*-ferulic acid (FA) in solubilized lignin from grass family biomass [7]. Despite the initial tolerance exhibited by the *P. putida* for aromatic compounds in general and to *p*-CA and FA in particular [8] there is still apparent toxicity when growing under high concentration of these acids derived from biodiverse lignin feedstocks [9]. Thus, innovative and more efficient routes need to be developed in order to utilize lignin for economically viable bioproduction processes.

Pseudomonas putida KT2440 (KT2440), is a promising conversion chassis for the bioprocesses utilizing lignocellulosic feedstocks. KT2440 is classified as biosafety level 1 [10] of *Pseudomonads* species which have a broad metabolic versatility and genetic flexibility [11,12] allowing their abundance in many different environments. Especially, they possess a capability to degrade many aromatic compounds [13] including the lignin-derived compounds. These advantages have made *P. putida* attractive for biotechnological applications including bioproduction of broad range of chemicals [14–16] and bioremediation of pollutants in the environment [15,17]. Although KT2440 inherently shows higher tolerance to the aromatic compounds when compared with *Escherichia coli*, probably due to membrane

integrity [8], high levels of these acids in lignocellulosic feedstocks still inhibit cellular growth and chemical production [9].

Adaptive laboratory evolution (ALE) is an efficient approach that has proven to be useful for overcoming toxicity of substrates and inhibitors in media and improving strain fitness [18–21]. In the present work, the toxicity problem of the aromatic acids is addressed by utilizing a systematic ALE strategy, called Tolerance Adaptive Laboratory Evolution (TALE) [18] to generate KT2440 strains capable of growing at elevated aromatic acid concentrations. To achieve this goal, three TALE experiments were designed; the cells were continuously evolved in independent biological replicates under increasing levels of *p*-CA, FA and a mixture of equal ratio of both acids mimicking a typical composition found in solubilized fractions from lignocellulosic feedstocks. In the meantime, ALE was applied to generate evolved KT2440 on glucose at a constant concentration for fast growth as previously described [19,20]. After the experiment completed, populations and clones were selected from both TALE and ALE endpoints as well as intermediate time points in order to identify causal mutations. Further, endpoint clones were characterized in terms of their growths rates, tolerance levels. It was found that causal mutations are related to outer membrane protein *algE* that is important for alginate secretion and efflux pump membrane protein *ttgB*. The generated platform strains as well as the findings could be utilized to develop the efficient biorefinery processes with maximizing potentials of *P. putida*.

Methods

Strain, medium, and reagent

P. putida strain KT2440 (ATCC 33849) gifted from Dr. Pablo I. Nikel (Novo Nordisk Foundation Center for Biosustainability) was utilized for this study. A modified M9 minimal medium (Ammonium sulphate instead of ammonium chloride) was used as a base medium for the all ALE experiments and characterization assays. The utilized M9 medium contained 1X M9 salts, 2 mM $MgSO_4$, 100 µM $CaCl_2$, 1X trace elements, and each carbon source, i.e., *p*-CA, FA, or a 1:1 mixture of, *p*-CA, FA, or glucose, dissolved in autoclaved *milli*-Q water. The 10X M9 salts consisted of 68 g/L Na_2HPO_4 anhydrous, 30 g/L KH_2PO_4, 5 g/L NaCl, and 20 g/L $(NH_4)_2SO_4$. Trace elements (2000X) consisted of 3 g/L $FeSO_4 \cdot 7H_2O$, 4.5 g/L $ZnSO_4 \cdot 7H_2O$, 0.3 g/L $CoCl_2 \cdot 6H_2O$, 0.4 g/L $Na_2MoO_4 \cdot 2H_2O$, 4.5 g/L $CaCl_2 \cdot H_2O$, 0.2 g/L $CuSO_4 \cdot 2H_2O$, 1 g/L H_3BO_3, 15 g/L disodium ethylenediaminetetraacetate, 0.1 g/L KI, 0.7 g/L $MnCl_2.4H_2O$. pH of the trace element solution was adjusted with hydrochloric acid solution to pH 4. Concentrated stocks of both *p*-CA and *trans*-FA (TCI-Europe N.v, Belgium) were prepared by dissolving each acid in water and neutralization with 5 M NaOH or directly using pellets when preparing large volume batch (pH 7). It should be noted that the titration was performed gradually and slowly in order to avoid degradation of the acids during pH shift and concentrations of *p*-CA and FA indicate amount of the acids, not including their sodium salt. Fresh batches of the stocks were prepared when needed to avoid degradation due to their short shelf-times during storage. Medium was sterilized by filtration with a 0.22-µm filter after mixing all ingredients.

For the routine propagation of cells and preparation of frozen stocks, cells were cultured in liquid or on agar plates containing Luria-Bertani medium (LB). When needed, antibiotics were used at the following concentrations: chloramphenicol (Cm) 30 µg/mL, kanamycin (Km) 50 µg/mL, and gentamicin (Gm) 10 µg/mL. All primers are listed in Supplementary file 1: Table S3.

Plasmid and strain construction

Deletion of PP_3350 was performed using the antibiotic/sacB method of gene replacement as described previously [3,22]. The plasmid pCJ222 for deleting PP_3350 was constructed by amplifying an upstream targeting region using a primer pair of oCJ791 and oCJ792 and a downstream targeting region using a primer pair of oCJ793 and oCJ794 from KT2440 genomic DNA. Then, these amplified fragments were assembled into the pK18sB plasmid digested with EcoRI and HindIII using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs, the United States). The assembly was transformed into NEB 5-alpha *F'Iq* cells. A clone with the correct sequence was identified using Sanger sequencing (GENEWIZ, Germany) with primers oCJ680 and oCJ681. pCJ222 was introduced into KT2440 as previously described [23] and sucrose selection and diagnostic colony PCR were followed to identify a clone with proper deletion mutation showing a 1097-bp amplicon when a primer pair of oCJ795 and oCJ796 were used for the PCR.

ttgB knock-out mutants were constructed using the plasmid pTE289. This plasmid was constructed using the isothermal DNA assembly method, as described in Gibson et al., 2009 [24]. The plasmid also contains the kanamycin resistance gene and the *sacB* gene, supplementary text File 1: Figure S12. The plasmid was introduced into KT2440 via conjugation with the *E. coli* S17 as a donor strain or into evolved KT2440 strains through electroporation. Then, the transformed cells were plated on LB agar plates containing kanamycin for selection of colonies with a single crossover of the target region. After 24 hours of growth at 30°C, single colonies were picked and grown in 5 mL LB media for 5 hours at 30°C and then plated in LB (no NaCl added) agar plates containing 25% (w/v) sucrose and grown overnight at 30°C for the counter-selection to obtain double crossover colonies not sensitive to sucrose. The colonies were further checked for confirmation of the correct double crossover via colony PCR with primers: TEAM-1640, TEAM-1641, F-5'ttgAend, and R-ttgCbeg, Supplementary File 1: Table S3. The correct double crossover resulted in 622-bp of amplicon when the last two primers were used for the colony PCR. The colonies were further streaked on pseudomonas isolation agar and pseudomonas isolation agar with kanamycin plates as a check. Colonies sensitive to kanamycin with the correct amplicon sizes were chosen. At least two isolates were grown in the liquid medium overnight and then stored in -80 °C.

Initial growth evaluation of the wildtype KT2440 tolerance

P. putida KT2440 was initially evaluated with different concentrations of each of the aromatic acids in order to choose proper starting concentrations where KT2440 does not shows severe decrease in its growth rate and final optical density for maintaining efficient passaging during the TALE. A detailed description of tolerance screening and tolerance phenotype in the wildtype KT2440 (Additional file 2: Table S1). Cells from an overnight culture in LB medium were inoculated into cultivation tubes containing 15 mL of the M9 medium supplemented with glucose and different concentrations of each aromatic acid. Inoculated culture tubes were stirred with a magnetic bar to 1,100 rpm at 30^{o} C.

Adaptive laboratory evolution (ALE)

The TALE and ALE experiments were systematically conducted using an automated liquid handler platform [18–20]. To prepare pre-cultures, a frozen KT2440 glycerol stock was streaked on a LB agar plate in order to obtain single colonies. Then, pre-cultures for inoculating the starting culture were aerobically grown in the M9 medium with glucose and 1% (v/v) of the pre-cultures was inoculated into 4 sets of tubes containing the same medium. The experiments were conducted with independent 6 replicates for the TALE and 4 replicates for the control glucose ALE in 15 mL of the M9 medium supplemented with the starting concentrations of the corresponding carbon sources (Table 1). During TALE or ALE, cells were aerobically cultivated in a heat block at 30 ^{o}C and passed once optical density at 600 nm (OD_{600}) reached approximately 1.3 ($OD_{600} \leq 2$). For the TALE experiments, when increased growth rate was achieved after a defined period of time at a particular concentration, acid concentration was increased. This process was repeated until a significant increase in tolerance was achieved. If there were occasions where the acid concentration increase resulted in severely ceased growth, the concentration was lowered to previous level in order to restore the growth. The ALE experiment was conducted similar to the TALE experiments except that constant concentration (10 g/L) of glucose was used as a sole carbon source. Periodically, culture samples were collected and frozen in 25% glycerol solution and stored at -80^{o}C for further need.

Table 1. Summary of the TALE and ALE experiments.

Condition	ALE #	Starting conc. (g/L)	Initial growth rate with starting conc.(h^{-1})	Average ending conc. (g/L)	Final growth rate with ending conc. (h^{-1}) *	Fold-change	Total number of flasks	Cumulative cell division (CCD), $x10^{12}$	Number of generations
Coumaric Acid	1	8	0.12 ±0.01	25.6	0.08	3.2	62	2.15	311.0
	3			27.2	0.17	3.4	69	2.28	347.5
	5			25.6	0.09	3.2	73	2.47	369.5
	7			27.2	0.08	3.4	70	2.35	352.6
	9			27.2	0.09	3.4	76	2.43	381.2
	11			25.6	0.12	3.2	66	2.18	336.0
				26 ± 0.80	0.10 ± 0.03				
Ferulic Acid	13	10	0.13 ± 0.01	40	0.26	5.0	104	3.09	574.3
	15			40	0.29	5.0	115	3.49	622.5
	17			40	0.34	5.0	108	3.42	594.8
	19			40	0.34	5.0	120	3.39	658.4
	21			40	0.34	5.0	114	3.33	622.8
	23			40	0.31	5.0	115	3.49	632.5
				40	0.3 ± 0.1				
Feru. & Coum. Mix (1:1)	25	6	0.18 ± 0.01	20.4	0.12	2.6	98	3.42	602.4
	27			20.4	0.11	2.6	104	3.63	538.6
	29			20.4	0.08	2.6	104	3.44	535.0
	31			20.4	0.12	2.6	103	3.51	531.5
	33			17.5	0.18	2.2	109	3.69	558.7
	35			19.8	0.18	2.5	97	3.16	493.3
				19.8 ± 1.1	0.13±0.036				
Glucose	37	10 (constant)	0.50 ± 0.01	10 (constant)	0.83	N/A	141	6.36	917.8
	38				0.85		143	6.53	934.9
	39				0.87		141	6.37	919.5
	40				0.83 ± 0.03		140	6.43	913.5

**The initial and final growth rates were calculated for ALE populations from the first and last 3 flasks, respectively.*

Re-sequencing of selected clones and populations

For whole genome re-sequencing (WGS), each endpoint population or clone from the 6 replicates of 3 different TALE experiments (*p*-CA, FA, and the mixture) and 4 replicates of the control ALE were selected. Additionally, 2 or 1 intermediate(s) from each of the TALE and ALE experiments, respectively was also included. In total, 44 and 45 populations and isolates were re-sequenced, respectively. Especially, each clone was chosen from an agar plate containing the M9 medium with a corresponding carbon source. Genomic DNA samples were extracted from their overnight cultures in LB medium at the stationary phase using the PureLink® genomic DNA extraction kit (Invitrogen, CA, the United States). the quality of the extracted DNA was initially was assessed by measuring relative absorbance at A260/A280 using Nanodrop spectrophotometer. Extracted DNA concentration was quantified using Qubit ds-DNA high sensitivity assay. Paired-end re-sequencing libraries were generated using the 300 cycle (150 bp x 2) kit from Illumina (San Diego, CA, the United States). Sequencing was performed on an Illumina Nextseq Sequencer. The average coverage for each of the re-sequenced sample was over 25X. Mutations were identified using a customized script as described by Phaneuf et al. [25] based on bowties2 [26] and Breseq (version 0.30.1) [27] with a reference genome of the wildtype KT2440 strain (accession number: NC_002947). When the population samples were analyzed, a filter was applied to exclude mutations with frequency less than 50% except for a mutation found in a clone derived from this population. Almost gene annotations or functions were retrieved from Pseudomonas Genome DB (http://pseudomonas.com) [28] or the Pfam protein families database [29] through homology comparison of sequences and functions to other *Pseudomonas* strains or other bacteria.

Growth evaluation of the endpoint clones and knockouts

Each of *p*-CA and FA acids were screened for their impact on bacterial growth, i.e. growth rate, final OD and lag time, using Bioscreen C^{TM} (Lab Systems Helsinki, Finland) screening platform as well as using cultivation tubes. Detailed description of screening methods and the results using Bioscreen C system are presented in the supplementary Text File: **Text S1.** Growth rates of the evolved clones were evaluated in the cultivation tubes. Cells were revived

from glycerol stocks on M9 minimal medium agar plates containing the substrates on which each strain was evolved as the sole carbon sources (5 g/L for *p*-CA, FA and the mix, respectively). Overnight cultures were performed in the cylindrical tubes containing 15 mL medium at 30 °C and 1100 rpm supplemented with the same concentration used in the M9 agar plates until OD_{600} reached approximately 3. Then, cells were inoculated at a low OD_{600} (1:100 dilution) in M9 medium supplemented with the specified carbon source for each culture condition. Inoculated tubes were temperature-controlled at 30 ^{o}C and fully aerated using a magnetic bar. Cultures were performed in duplicate with the same method used for TALE or ALE. The wildtype KT2440 was also revived in the M9 medium containing the corresponding carbon source (e.g., 3 g/L *p*-CA if further characterized with *p*-CA), cultivated in the same conditions as the evolved strains, and used as a control. Total 18 strains, i.e. 17 endpoint clones and the wild type control strain were evaluated in 4 different culture conditions with varied carbon sources and their concentrations: (1) 10 and 20 g/L *p*-CA, (2) 10 and 30 g/L FA, (3) 10 and 20 g/L the mixture, and (4) 10 g/L glucose. Growth rate and OD_{600} were measured using the Sunrise plate reader (Tecan, Männedorf, Switzerland). Results present average values from biological duplicates. Error bars show the absolute difference between the two replicates.

Similarly, Growth evaluation of the ΔttgB mutant and the ALE strains #25, 27, 29, and 31 were performed in the following manner. The strains were revived in Pseudomonas isolation agar grown overnight (O/N) at 30°C. Single isolated colonies were picked and grown in 5 mL of the modified M9 minimal media with selected carbon source (coumarate, ferulate, and coumarate and ferulate mix) at 30°C ON. These colonies were then grown overnight for a second time to allow for the cells to briefly adapt and condition to the new environment. Conditioned cultures were then diluted 1:100 to an OD_{600} 0.01-0.02 in growth medium and grown in 48-well flower plates in Biolector®. Light scattering measurements were collected every 15 minutes time intervals and converted to optical density OD_{600nm} using standard curve calibration.

Results

A Tolerance Adaptive Laboratory Evolution (TALE) experiment was utilized to generate strains of *Pseudomonas putida* KT2440 (KT2440) which could tolerate high concentrations of the aromatic acids, *p*-coumarate and ferulate. Additionally, a control ALE experiment was performed with glucose as the sole carbon and energy source as a comparative control. The two processes of ALE and TALE were successful in generating strains with either increased tolerance to the three aromatic acids conditions; *p*-CA, FA and a 1:1 mix of *p*-CA and FA or improved fitness, i.e. growth rate., for the control experiment on glucose. The TALE and ALE experiments were performed with six and four independently evolved replicates for each condition, respectively. For the TALE experiments, the aromatic acid concentration was increased under continuous exponential batch cultivation over the course of each experiment when growth rate reached a threshold of ≥ 0.15 hr^{-1}(see Methods for more details). Overall, there were (3-4)-fold increases in aromatic acid concentrations for each TALE experiment over the starting concentration (Table 1). The observed growth rate trajectories of representative replicates from each of the TALE conditions, as well as the ALE experiment, are presented in Figure 1. Similar plots for the remaining populations from all conditions are shown in supplementary text File 1: Figure S5 and Figure S6. It is important to note that there was a jump in the fitness across all TALE experiments just after the experiment start as the aromatic acids' concentrations were incorrectly decreased due to an incorrect parameter in the process control setup, instead of increasing – this was corrected shortly after the experiment started.

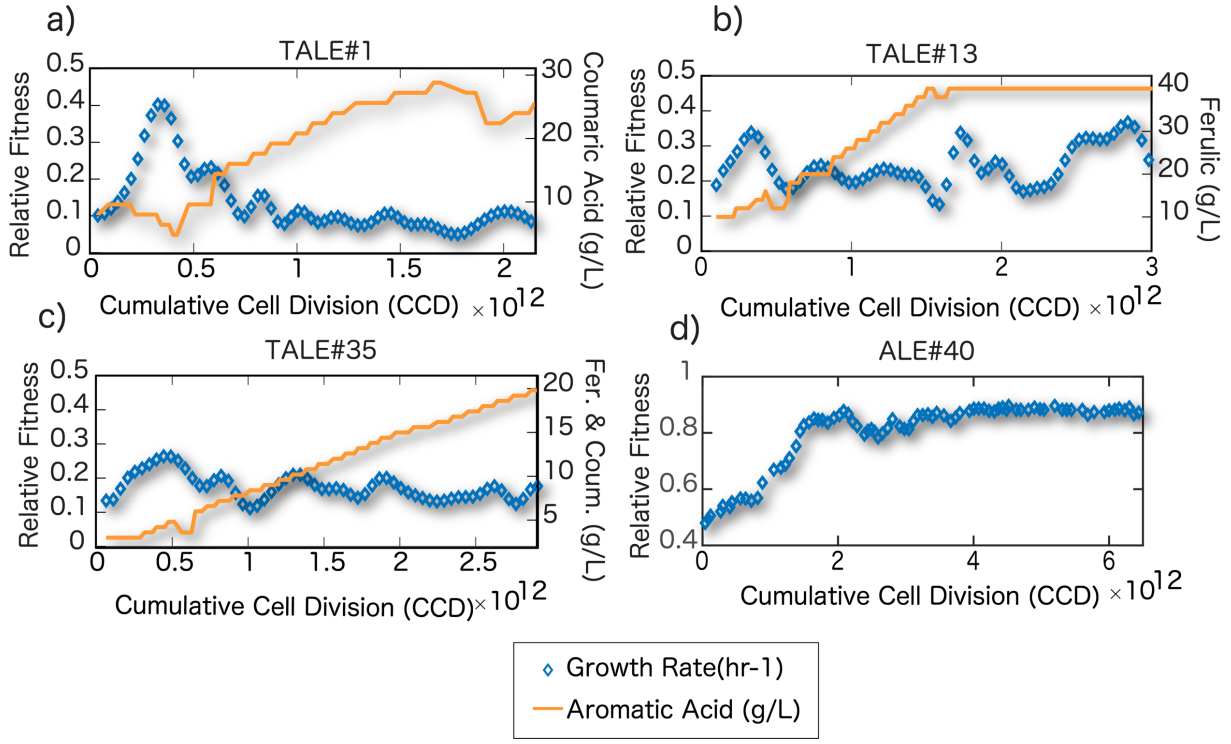


Figure 1. Representative trajectories of population fitness and aromatic acid concentrations over the courses of 4 evolution experiments. Depicted are the fitness trajectories vs the dynamic change of aromatic acids concentration over the course of the representing TALE experiments along with representative ALE on glucose versus the cumulative cells division (CCD) experienced by a) TALE#1, b) TALE#13, c) TALE#35 and ALE#40 for *p*-CA, FA, the mixture and glucose, respectively. Concentration of each aromatic acid was shown as an orange line and growth rate was shown as a blue diamond.

The ability of the KT2440 cells to adapt to increasing concentration of each aromatic acid varied with respect to the condition. For each of the TALE conditions of *p*-CA, FA and the mix, the final average aromatic acids concentrations achieved was 26±0.80, 40, and 19.8±1.1 g/L, respectively. Under the ALE condition with glucose as sole carbon source, each population exhibited an increase in growth rate relative to the starting strain Table 1. The growth rate increase was 1.69-fold faster than the starting strain. Overall, there was only one fitness, i.e. growth rate, jump observed across all independent glucose ALE replicates. Each experiment completed with an average of 69, 112, 102, 141 flasks, i.e. passages or transfers, for conditions of *p*-CA, FA, mix (*p*-CA&FA) and glucose, respectively. Overall, each of the TALE conditions of *p*-CA, FA and a 1:1 mixture of *p*-CA and FA and ALE on glucose populations underwent an average total number of cumulative cell divisions (CCD) of $2.31x10^{12}$±0.12, $3.36x10^{12}$± 0.13, $3.47x10^{12}$±0.17 and $6.42x10^{12}$±0.06, respectively. The use of CCD was previously demonstrated as a very meaningful timescale for ALE experiments [30]. Sequencing and screening of the evolved endpoints isolates from across different conditions was performed in order to understand overall performance and link their phenotype to their respective genotype.

Whole genome resequencing and mutation analysis

Whole genome resequencing was used to determine the genetic basis for the increased tolerance phenotypes. Key mutations were determined by comparing all clonal or population samples from both endpoint and intermediate flasks and identifying genes or genetic regions (i.e., intergenic regions) which had multiple unique mutations or were mutated across isolates from independent replicates. Key mutations are listed in Table 2 and the full list of mutations is in Supplementary data file 3. Overall, there were 62, 147, 78 and 28 unique mutations identified for *p*-CA, Ferulic, mix of *p*-CA and FA and Glucose conditions, respectively, across all re-sequenced clones and populations. Interestingly, there were no obvious hyper-mutating linages identified across the different evolution experiments.

Table 2. The key mutations from the TALE and ALE experiments. The mutations were listed if a given gene or genetic region were mutated several times in a unique way or identical mutation occurred multiple times across independent replicates. Abbreviations: SNP, single nucleotide polymorphism; INS, insertion; DEL, deletion; DUP, duplication; CA, *p*-coumarate; FA, *trans*-ferulate; Mix, the mixture; Gluc, glucose.

Mutation type	Gene / Genetic Region	Unique Occurrences	Mutation types (count unique)	Gene Function	Condition
Shared	*gacS*	17	SNP (7), DEL (3)	sensor protein GacS	CA, FA, mix, Gluc
	oprB-II	7	SNP (2), DEL (1)	carbohydrate-selective porin	Mix, Gluc
	uvrY	8	SNP (6), INS (1)	BarA/UvrY two-component system response regulator	CA, FA, Mix, Gluc
	sucA	12	SNP (7)	2-oxoglutarate dehydrogenase subunit E1	CA, FA, mix
	rho	17	SNP (10)	transcription termination factor Rho	CA, FA, mix
	PP_3350	26	SNP (6), DEL (12), INS/DUP (3), INS (1)	hypothetical protein	CA, FA, mix
	ttgB	14	SNP (6)	efflux pump membrane protein TtgB	CA, Mix
	accD	6	SNP (3)	acetyl-CoA carboxylase carboxyltransferase subunit beta	CA, FA, Mix
	PP_5245, *kefB*-III	15	SNP (2)	AraC family transcriptional regulator/glutathione-regulated potassium/H+ antiporter	CA, FA, Mix
	fleQ	6	SNP (6), DEL (1)	transcriptional regulator FleQ	FA, Mix, Gluc
Condition-specific	*accA*	2	SNP (2)	acetyl-CoA carboxylase carboxyltransferase subunit alpha	CA
	dnaK	2	SNP (2)	chaperone protein DnaK	FA
	wbpL	3	SNP (1), INS (2)	glycosyl transferase WbpL	CA
	PP_1060, PP_1061	2	SNP (2)	glutamate synthase large subunit/ATP-dependent DNA helicase	CA
	rpoN	3	INS (1), DEL (1)	RNA polymerase sigma-54 factor	Mix
	sdhA	1	SNP (1)	succinate dehydrogenase flavoprotein subunit	Mix
	oxyR	3	SNP (1), INS/DUP (1)	oxidative and nitrosative stress transcriptional dual regulator	Gluc
	fliR	1	DUP (1)	flagellar biosynthetic protein FliR	Gluc

The key mutations were grouped into two categories; “shared” mutations identified in across different TALE and ALE conditions and “condition-specific” mutations which were identified in only one substrate condition. Overall, there were 11 genetic regions were identified as “shared” mutations. Furthermore, there were three, two and two “condition-specific” key mutations for *p*-CA, FA and the mixture, respectively. Interestingly, all key mutation identified in the control ALE on glucose were shared across all TALE conditions (**Figure 2)**.

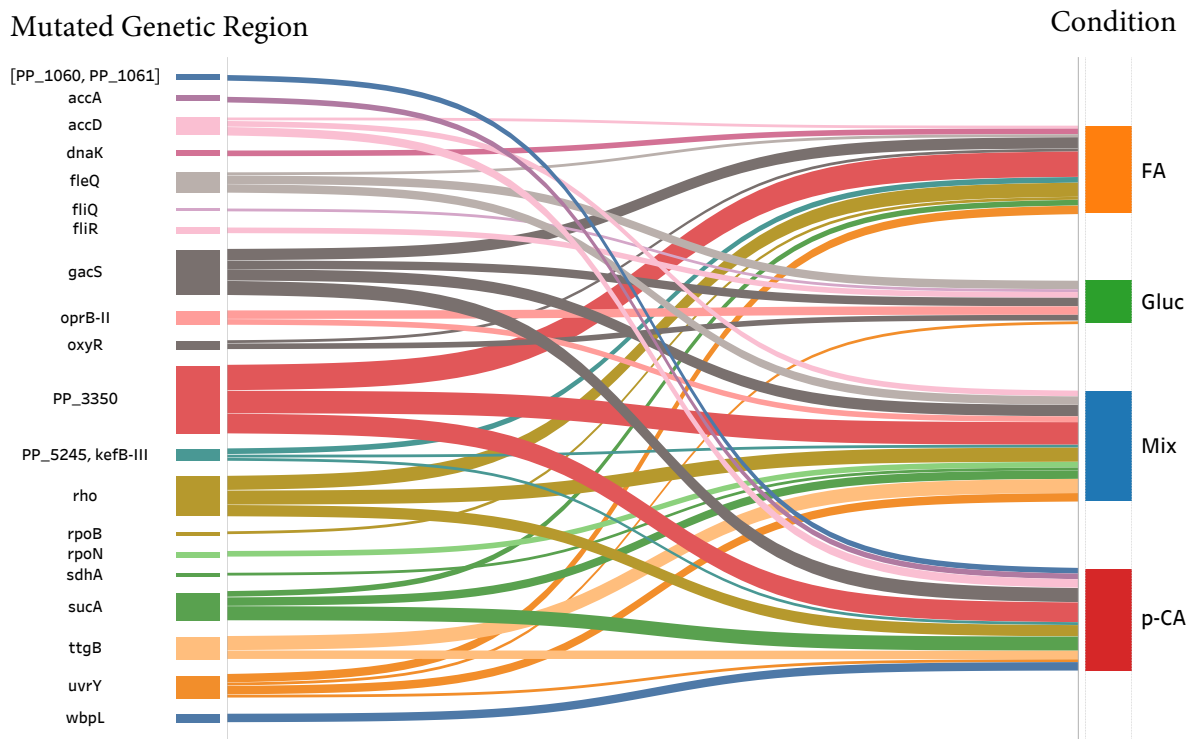


Figure 2. A landscape of the key mutations occurred across the 4 different evolution experiments. The graph shows an overview of the genes mutated in the different ALE conditions and the cross occurrences of several mutations in the same genomic region (see Table 2). The width of the lines is proportional to the number of unique occurrences in which a given gene or genetic region was mutated.

Two genes were identified in across all different TALE and ALE conditions were involved in metabolic regulation processes including gacS and uvrY genes. The *gacS* gene encodes for the GacS sensor kinase as a constituent of the sensor element of the GacS/GacA two-component system [31] and was proposed to influence the biofilm formation. A total of 7 different SNPs and three deletions were found in *gacS* and some of them were likely loss of function mutations which created a stop codon or were out of frame deletions. The role of all of these mutations in terms of tolerance was not immediately apparent. Extra shared mutations were found in the *uvrY* gene, which encodes for BarA/UvrY two-component system response regulator, all ALE conditions. This gene was mutated 6 times, 5 SNPs and one unique insertion mutation.

In addition, shared mutations were found in the carbohydrate selection porin, *oprB-II* [32]. Interestingly, these mutations were identified only in two conditions the control ALE on glucose with three independent occurrences and five, four and five mutations occurring in the *p*-CA, Ferulic and mixture condition of *p*-CA and FA, respectively, Table 2. OprB-II encodes for a glucose inducible protein (*OprB*) which was found to have glucose uptake affinity [32]. Mutations targeted this gene were a mixture of SNPs and deletions; three SNPs, two identical SNP mutations and the same one base pair deletion mutations observed across the two conditions.

The most predominant mutated gene across the tolerance ALE conditions, i.e. *p*-CA, FA and the mixture of *p*-CA and FA, was in the ORF annotated as a hypothetical protein, PP_3350 [29] with 26 unique mutation events. Surprisingly, no mutations in this gene were identified in the control ALE on glucose. This indicated that the mutation is likely adaptive for growth and tolerance on aromatic acids for *P. putida*. Mutations in PP_3350 were a group of different SNPs, deletions of different sizes, short deletions with ranging from 1 to 115 bp or long deletions of 61,130 bp or 69,784 bp containing 55-56 genes, and finally a short 31-bp duplication insertion. A number of these mutations suggest a potential loss of function of this specific gene. Given the high frequency of mutations occurrences in the PP_3350 gene, a follow up analysis was performed to assess its role on tolerance on the different acids (see the reverse engineering section).

Furthermore, investigation of PP_3350 homologous genes in closely related Pseudomonas species using Pfam database has been identified as *algE* gene and has 100%

homology with *algE* protein sequence(NCBI GenBank: HBK50584.1) [29,33]. AlgE is an outer membrane protein that is required for production of extracellular polysaccharide, alginate [34] and as a member of the general diffusion porin family. Additionally, algE has been characterized in i.e. *P. aeruginosa* and it was found that *algE* mutant was nonmucoid and produce no detectable levels of alginate [35]. Although, the mechanistic role of mutations in the tolerance phenotype is still not clear and needs further studies. However, plausible explanation for improved tolerance phenotype demonstrated by deleting the PP_3350 in the background strain KT2440 is either, preventing the alginate exopolysaccharide secretion and subsequently leads to alginate degradations then acts as energy saving [36] or prevent passive diffusion of small molecules such as aromatic acids.

There were several additional genes mutated in all three TALE experiment were involved in tricarboxylic acid cycle, transcription regulation, membrane transport and carbon metabolism. Among them, twelve mutation occurrences were detected in the *sucA* gene, predicted by homology with other Pseudomonads to encode for 2-oxoglutaric dehydrogenase subunit E1 [29]. There was a total of 7 unique SNP mutations observed in three ALE conditions, *p*-CA, FA and the mixture condition of both. In addition, other events of SNP mutations were identified in the *rho* gene which encodes for the transcription termination factor Rho [29]. This gene had around 17 instances of unique occurrences across different re-sequenced populations and clones. In total there were about 10 unique SNP mutations identified in the three TALE conditions, *p*-CA, FA and the mix of both acids. Another shared mutation in all TALE conditions, were two unique SNP mutations in the intergenic region of PP_5245 encoding the AraC family transcriptional regulator and the *kefB*-III encoding glutathione-regulated potassium/H+ antiporter with total occurrences of 15 times. Similarly, another identified gene, *accD* predicted as acetyl-CoA carboxylase carboxyltransferase subunit beta participating in the lipid biosynthesis [29]; 3 unique SNPs were found in this gene.

Additionally, one gene was mutated in glucose ALE, FA, the mixture condition of both *p*-CA and FA conditions; *fleQ* gene. The *fleQ* gene encoding for transcriptional regulator FleQ [29] had several mutations including two unique SNPs and a 2,866-bp deletion which included two upstream genes of *atoC* and *fleS*. These two genes encode two component system AtoC DNA-binding transcriptional activator and sensory box histidine kinase FleS, respectively [29]. Interestingly, deletion of *atoC* in *E. coli* enhanced the ability to utilize organic acids through overexpression of acetoacetyl-CoA transferase encoded by *atoDA*

[37]. The later structural deletion mutations suggest a loss of function. FleQ regulates a lot of putative genes including flagellar, motility-related genes, other genes involved in adhesion and exopolysaccharide production [38]. Mutations introduced into *fleQ* were found to reduce biofilm formation [39].

Another identified gene in two TALE conditions; the *p*-CA and the mixture conditions was involved in membrane transport. This gene was *ttgB* that was mutated 14 times with 6 unique SNPs. TtgB was predicted to encode for efflux pump membrane protein TtgB without known specificity [29]. TtgB is a member of the TtgABC efflux system and the expression of the efflux complex improved tolerance in *Pseudomonas* from a toxic short-chain alcohol [40] and toluene [41].

Condition-specific mutations observed in *p*-CA TALE, include genes involved in carbon metabolism. Three mutations were two insertion mutations and one SNP in *wbpL*, encoding for glycosyl transferase WbpL [29]. The insertion mutations were, one nucleotide insertion (thymine) and 7 nucleotides (guanine) duplication insertion, resulting in frameshift. Previous study showed that a *wbpL* knock-out mutant was found to be tolerant to organic solvents shocks [42]. Furthermore, as *p*-CA TALE specific mutations, two intergenic SNP mutations were identified between PP_1060, PP_1061 genes encoding glutamate synthase large subunit/ATP-dependent DNA helicase [29]. Finally, two coding SNPs in the *accA* gene that encodes for acetyl-CoA carboxylase carboxyltransferase subunit alpha as a component of the acetyl coenzyme A carboxylase (ACC) complex were identified only in one population replicate and the its subsequent clone [29]. The product of this gene and other genes *accD* and *accC* are predicted to be involved in the first step of fatty acid biosynthesis, which is important for cell membrane synthesis [29]. Modulation in cell membrane lipids was found to play an important role in tolerance phenotype of *Pseudomonas* strains for either toluene and vanillin [43,44].

For the ferulate TALE, two SNP mutations were identified in *dnaK* encoding for chaperone protein DnaK [29]. This gene is known to be one of several heat-shock inducible genes in *P. putida* [45]. This gene was found, along with other chaperonins, to be upregulated when *P. putida* strain was exposed to toluene. It was studied that these chaperonins play important rule in protein-damage repair system [46] during solvent exposure. Moreover,

Upregulation of *DnaK* was observed as one of butanol tolerance response in a *P. putida* mutant [47].

Condition-specific mutation in the mixture of TALE isolates included gene responsible for transcriptional activation *rpoN*. RpoN, encoding RNA polymerase sigma-54 factor had one +A insertion mutation and another 11-bp deletion mutation. RpoN is a part of the two-component system for class III flagellar system responsible for flagellar assembly. Similarly, additional gene was mutated only one time, *sdhA* in the mixture conditions. SdhA encodes for succinate dehydrogenase flavoprotein subunit.

Further, two genes were identified only in the glucose ALE condition including *oxyR* and *fliR*. The first gene was mutations of *oxyR* that encodes for oxidative and nitrosative stress transcriptional dual regulator [29]. This gene was mutated three times with one unique SNP mutation and two insertion duplication short sequence. Previous study found that *oxyR* mutant had increased the amount of several antioxidative stress proteins beside its role in peroxide protection on *P. putida* KT2442 strain [48]. The *fliR* gene encoding for flagellar biosynthetic protein FliR [29]. This gene was mutated one time with the same insertion duplication sequence across glucose ALE conditions.

Characterization of selected endpoint clones

Phenotypic growth screens for a total of 17 clones (I2, supplementary data File 1) isolated from each endpoint populations revealed improved tolerance and cross-tolerance to aromatic acids or fitness increases in the glucose minimal media condition. All isolates were capable of growth in M9 minimal medium in the presence of four different conditions (substrate); *p*-CA, FA, a mixture of (1:1) *p*-CA and FA (mix) or glucose as the sole carbon source. The improvements in performance of the isolates were compared to the control strain (wildtype KT2440). In order to quantify the improved tolerance or growth on a given substrate, growth rate was used as an indicator. Examining the phenotypes of the evolved clones for tolerance of *p*-CA revealed that at low concentrations, there were few clones with improved performance and at high concentrations there was a more marked improvement, specifically from clones evolved on *p*-CA and the mixture of *p*-CA and FA conditions. Endpoint clones were characterized with *p*-CA at low (10 g/L, Figure S10A) and high concentrations (20 g/L, Figure 3A).

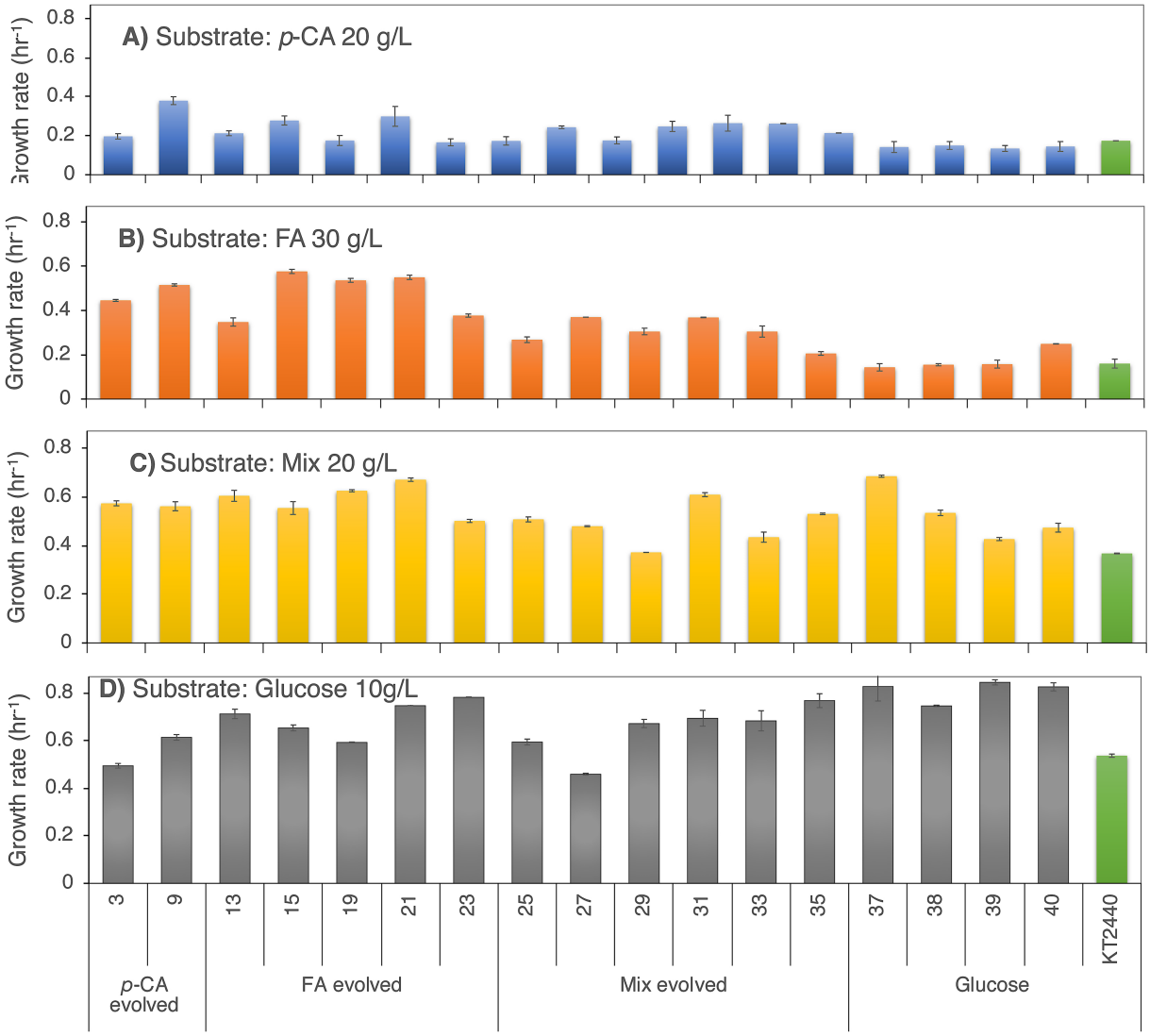


Figure 3. Screening of the endpoint clones in different substrates and at different concentrations vs the wildtype KT2440 strain. Depicted are growth rates for each endpoint clone when evaluated with A) *p*-CA (20 g/L), B) FA (30 g/L), C) the mixture (20 g/L, 10+10 g/L) and glucose (10 g/L). The wildtype KT2440 strain, shown in green, was used as the control across different testing conditions. Bar indicates growth rate (hr^{-1})

At 10 g/L *p*-CA (Figure S10A), all strains were capable of growth and 7 of the 17 ALE endpoint isolates had demonstrated higher growth rates compared to the growth of the control strain (see Supplementary data File 2). The highest growth rates were observed for strains in the *p*-CA, FA, and the mix condition groups with growth rates of up to 0.45 hr^{-1} for a *p*-CA evolved clone (lineage 9). However, three isolates, one from each of the *p*-CA, FA, and the mix *p*-CA+FA evolved linages demonstrated lower growth rates than of the control. At 20 g/L *p*-CA, while all strains were capable of growth at this high concentration, only 6 TALE evolved isolates demonstrated significantly increased growth rates over the control strain, whereas there were no clones that had a significant decrease. Isolates from the ALE experiment on glucose showed a similar fitness as the control (with an overall average growth rate slightly lower). Overall, there were 1 out of 2 isolates from the *p*-CA condition TALE which shown the most improved isolate and significantly displayed a higher growth rate at the high concentration of *p*-CA and the highest growth rate of any clone was 0.37 hr^{-1} and was again for the endpoint clone from linage 9. Looking at individual isolates, the endpoint clones of lineages 9, 23 and 19 (growth rate of 0.45, 0.42 and 0.40 hr-1, respectively) may be attractive for processes using either low or high *p*-CA concentrations (0.37, 0.28 and 0.30 hr-1, respectively) as they had demonstrated higher fitness at both conditions compared to the control strain (0.30 and 0.17 hr-1, respectively).

Similarly, examining the growth phenotype of the evolved clones for tolerance on FA revealed that there were few clones with improved performance at low concentration and at high concentration there were a noticeable improve in performance, significantly from clones evolved on FA. Endpoint clones from the independent lineages were also characterized under low (10 g/L, Figure S10D) and high (30 g/L, Figure 3B) FA concentrations. At 10 g/L FA, all evolved isolates were able to grow in the presence of the low FA concentration. Moreover, 2 out the of the 17 isolates showed higher growth rates (growth rate of 0.59 hr^{-1} and 0.57 hr^{-1} for lineages 31 and 27, respectively) when compared to the control strain (0.53 hr^{-1}) both were derived from the mix TALE. However, there were five isolates, two from p-CA TALE and two from the mixture TALE, that showed lower growth rates than the control strain (see Supplementary data File 2). At the highest FA concentration (30 g/L), all evolved isolates were capable of growing (Figure 3B). Further, all of the TALE-evolved strains had higher growth rates (growth rate between 0.47 hr^{-1} and 0.57 hr^{-1}) than the control wild type strain (0.16 hr^{-1}).

Overall, the highest growth rates were observed for aromatic acids evolved strains. Of the examined isolates, clones from lineages 3, 9, 15, 19 and 21 (two evolved for *p*-CA and three evolved for FA, respectively) showed the highest growth rates (> 0.35 hr^{-1}).

The third testing condition was the mixture of *p*-CA and FA. Concentrations were 10 g/L total aromatic acids (5 g/L *p*-CA + 5 g/L FA) (Figure S10F) and 20 g/L total aromatic acids (10 g/L p-CA + 10 g/L FA) (Figure 3C). There were clones showed high similarities in performance with the two condition; clones 20 and 21 clearly showed increased growth rates over the control at both low and high concentrations. There was no evolved clone which had a significantly lower growth rate than the control. Most of the evolved strains, including the evolved strains on glucose, demonstrated higher growth rates (up to 0.68 hr^{-1}) than the control (0.34 and 0.36 hr^{-1}, respectively). Some of the best performing isolates from different lineages which had improved growth rates (above 0.53 hr^{-1}) included isolates 13, 15, 19, 21, 31 and 37. These clones were from the FA, the mixture, and glucose ALE conditions.

For the control condition on glucose, all evolved strains were capable of growth on 10 g/L glucose. Surprisingly nearly all (15 out of 17) strains demonstrated higher growth rates (up to 0.84 hr^{-1} for lineage #39) regardless of the ALE conditions, compared to the control strain (0.53 hr^{-1}). Especially, all of the clones from the glucose ALE showed increased growth rates, with a 1.5-fold increase in average, over the wildtype growth rate. Two evolved clones had a lower growth rate than the control on glucose, these were from lineages 3 and 27 (evolved for *p*-CA and the mixture, respectively).

Causality validation of two key genes: PP_3350 and *ttgB*

Two of the most frequently mutated genes in this tolerance ALE study were the PP_3350 ORF, a hypothetical protein, and the *ttgB* gene, a subunit of a membrane efflux pump (Table 2). To examine their causality, gene deletion (i.e., knockout) studies were performed on both the wild type starting strain and several representative evolved populations. The findings yielded inverse results for the two genes which implicates their roles but establishes each as containing causal mutations. Each will be addressed.

To examine the causality of the PP_3350 open reading frame (ORF) mutated several times in this study, growth screens were conducted utilizing a ∆pp3350 strain which was benchmarked alongside a group of selected evolved isolates and wild type. The selected evolved endpoints were from populations 7, 23 and 25 (I0R2, supplementary data File 1), which were evolved on *p*-CA, FA, and mixture of (*p*-CA + FA), respectively. These respective strains harbored mutations in the PP_3350 gene and showed improved performance compared to the starting wild type KT2440. Overall, the PP_3350 genes was confirmed to play an important role in aromatic acids tolerance and utilization since its deletion showed phenotypic changes compared to the control strain on each of the three acid substrate conditions – most of the times showing a growth rate advantage.

The PP_3350 gene deletion had a positive impact on growth on coumarate at both high and low concentrations of the acid. The ∆pp3350 strain showed a growth rate (0.43 hr^{-1}) comparable to the evolved strains and higher than the wild-type strain (0.30 hr^{-1}) when grown at a lower *p*-CA concentration (Figure 4A). Moreover, lag times were decreased (≤ 5 hrs) compared to the control (11.7 hrs) and were similar to the lag times of the selected evolved populations. At a higher *p*-CA concentration, the ∆pp3350 strain also performed similarly to the evolved strains with significantly increased growth rates and decreased lag times compared to the wild-type strain. On the other hand, the PP_3350 deletion had no impact on growth at low ferulate concentration compared to the wild-type strain (Figure 4B). At this concentration, all strains performed the same as the wild-type strain. However, at the high FA concentration, the PP_3350 deletion had a positive impact on growth. This impact was manifested both as increased growth rate (0.16 hr^{-1} vs 0.11 hr^{-1}) and as shorter lag time (7.7 vs 9.9 hrs). These values were broadly similar to those obtained for evolved. Some of the evolved strains (specially the one evolved in FA, TALE 23) had better performance than ∆pp3350 (growth rate of 0.26 hr^{-1} and lag time of 5 hrs) suggesting that there are other mutations that provide a growth benefit in FA beyond the PP_3350 deletion. Finally, the PP_3350 deletion had showed negative impact on the growth at lower concentration and positive impact at higher concentration compared on the control strain. The ∆pp3350 strain exhibited growth disadvantage compared to the control strain but at the same level as the evolved strains in terms of growth rate and lag time (growth rate of 0.35 hr^{-1} vs 0.48 hr^{-1} and lag time of 4.0 hrs vs 4.7 hrs, respectively) Figure 4C. On the other hand, deleting PP_3350

seems to provide a slight growth advantage compared to control strain when using high concentrations of the mix as substrate (growth rate 0.22 hr^{-1} vs 0.16 hr^{-1}, respectively) with much shorter lag time (6 hrs vs 11 hrs, respectively) similar to evolved strains Figure 4C. The PP_3350 deletion had demonstrated a similar growth advantage in the control strain comparable to the evolved strains.

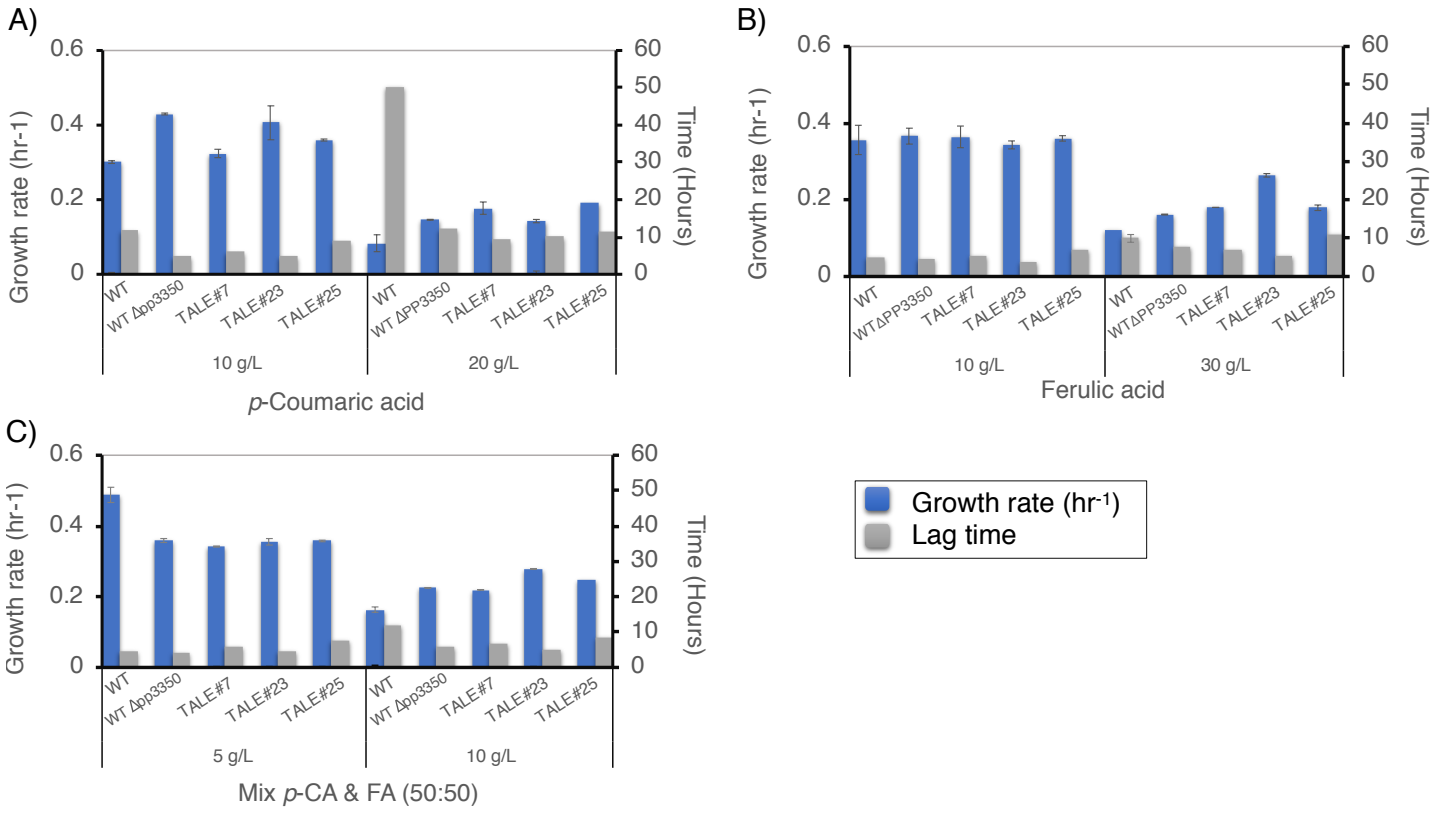


Figure 4. Comparison of the growth rates and the lag times of the wildtype, Δpp3350 and selected evolved strains in different growth conditions. The strains were grown in M9 medium containing different concentrations of A) *p*-CA, B) FA and C) the mixture as sole carbon sources. Left blue bar indicates growth rate (hr^{-1}) and right grey bar indicates lag time (Hours).

To examine the causality of the *ttgB* gene that was mutated several times in this study, especially in the *p*-CA TALE and the mix conditions, growth screens were conducted utilizing a Δ*ttgB* strain which was benchmarked alongside a group of selected evolved endpoints and wild type. The selected evolved strains were isolates number one (i.e., 'I1', supplementary data File 1) derived from endpoint populations 25, 27, 29 and 31 which were evolved on the mixture of (*p*-CA + FA). Overall, the *ttgB* gene was confirmed to play a significant role in aromatic acid tolerance and utilization since its deletion in the evolved strains showed detrimental phenotypic changes on either low or high concentrations of the mixture condition Figure 5. The negative impact of *ttgB* deletion was insignificant on the growth of the control strain KT 2440 compared to TALE derived strains. All evolved strains experienced a significant decrease in growth rate when the *ttgB* was knocked out. At low concentration (5 g/L), all evolved strains showed a growth rate of (equal or lower than 0.10 hr^{-1}) compared to 0.31 hr^{-1} for the control strain. At high concentration (10 g/L), all evolved strains performance was very poor with an overall growth rate ranging from growth rate less than 0.017 hr^{-1} (population 27) to very low to no detectable growth (population #25, 29, 31).

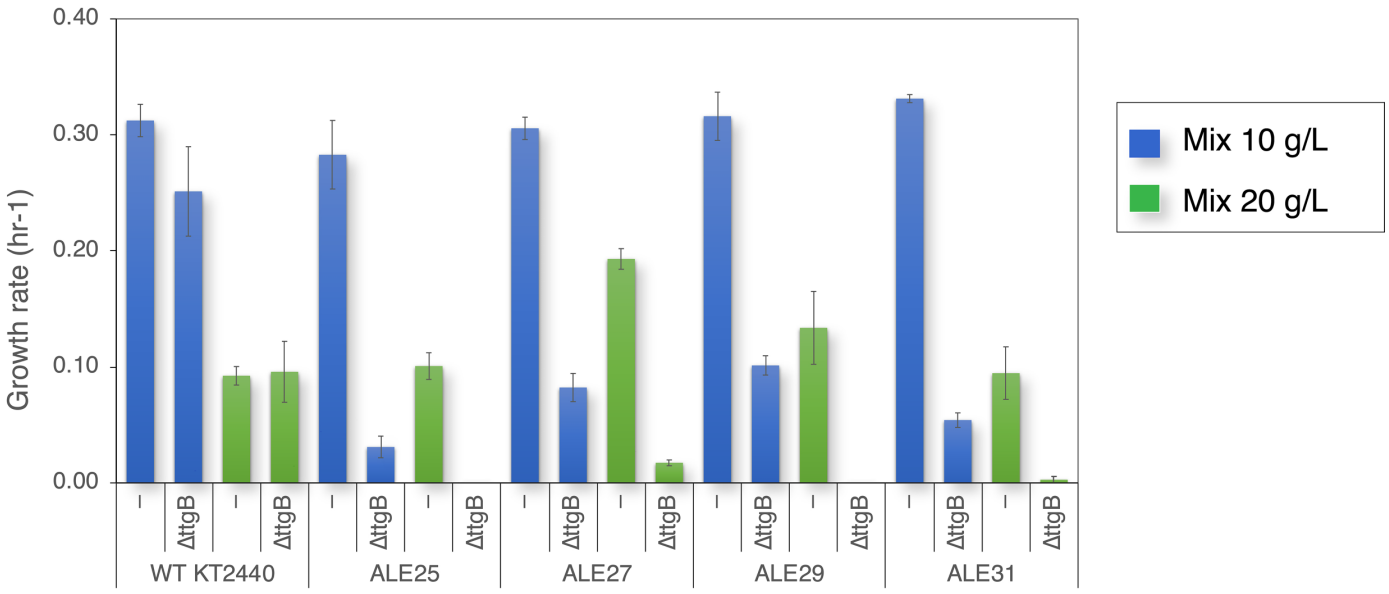


Figure 5. Comparison of the growth rates of the wildtype strain, the selected evolved strains and their Δttg B mutants with the mixture. Depicted are the growth rates of the wildtype KT2440 strain and the selected evolved strains from lineages, lineage#25 (ALE 25), lineage#27 (ALE 27), lineage#31 (ALE 31) as well as their *ttgB* knocked-out mutants. Blue bar indicates the results with 10 g/L the mixture and green bar indicates the results with 20 g/L the mixture.

Discussion

Utilization of biomass for microbial cell factories have several advantages due to the cheap prices and renewable, abundant sources which can help to establish feasible bio-production processes. Lignocellulosic biomass, with its large scale of abundance, represents an attractive raw feedstock. Further, when deconstructing lignocellulosic biomass, the lysate products typically have a surplus of monomeric bricks of aromatic acids. However, the utilization of these aromatic acids, such as *p*-CA and *trans*-FA, can be hindered due to their high concentration that makes it toxic even to native consumers. Thus, the scope of the current work is to generate aromatic acid tolerant strains through the adaptive laboratory evolution (ALE) approach for the efficient utilization. Accordingly, the main contributions from this work are: (1) effective and rapid generation of aromatic acids tolerant platform strains. This includes the cross-tolerant strains using an automated ALE platform for two common aromatic acids which can be used as platform strains for the utilization of feedstocks from either industrial waste or lignocellulosic degradation rich with the two carbon sources (*p*-coumarate and ferulate). Further, ALE was successful in generation of fitness improved strains on glucose, (2) revealing fitness and tolerance enabled mutations through systematic resequencing of isolates from multiple populations and (3) benchmarking key mutations discovered through gene knockout and characterizing their performance under different stress conditions

The ALE approach used in this study was successful in generating strains with either improved tolerance to the presence of aromatic acids or improved fitness on glucose compared to the wild-type strain over the course of ALE experiments under a well-controlled continuous passing routine during the exponential phase of growth. At the end of ALE, population replicates could grow at concentrations approximately 3.3±0.14, 3.3±0.17 folds higher than the starting strain for *p*-coumarate, ferulate and a mixture (1:1) of both, respectively (Table 1). Similarly, populations developed by ALE on glucose were able to grow approximately 1.69 ± 0.033-fold faster than the starting strain. This was further validated by characterizing endpoint clones in order to choose the most promising candidates for different substrate tolerance and utilization. Moreover, characterized endpoint isolates demonstrated

a high level of cross-tolerance at different conditions. For example, endpoint clones of lineages 3 and 7 derived from *p*-CA TALE demonstrated improved tolerance towards high concentration of FA (30 g/L), and endpoints lineages 15 and 19 (derived from FA) demonstrated higher fitness at both low (10 g/L) and high (20 g/L) p-CA concentration compared to the control strain (see results). Furthermore, almost all of the characterized endpoint isolates demonstrated improved tolerance at mixture conditions (at both low and high concentration) and improved fitness at glucose 10 g/L. Typically, in a given feedstock of solubilized lignin, *p*-ca and FA both coexist along with glucose [7]. Thus, TALE-derived strains with cross-tolerance to both aromatic acids high glucose consumption will have great advantages for biomass utilization.

As demonstrated in this study, using the automated ALE platform with multiple independent replicates under different conditions, combined with whole genome re-sequencing, has allowed deciphering the novel adaptive mutations through the identification of the key occurring mutations (**Figure 2**). This approach was strongly effective in identifying the key causal mutations via investigating mutated genes and regions several times across different conditions. The analysis revealed that the key mutations are involved in metabolic regulatory, transportation, and termination processes; those genes were *ttgB*, *oprB*-II, PP_3350 (*algE*), *rho*, *gacS*, *accA*, *accD, sucA, uvrY* and the intergenic region between PP_5245/*kefB*-III. Furthermore, the growth screen of two of the key mutations in genes PP_3350 and *ttgB* has confirmed their causality. The results from the growth screen of knocking out PP_3350 revealed that the gene is beneficial in a medium containing *p*-coumarate and that the growth advantage of this unique mutation is more significant at higher concentrations of aromatic compounds (Supplementary text File 2: Figure S11). Deletion of the *ttgB* gene has revealed that the presence of the *ttgB* gene is important for aromatic acid tolerance. Deleting *ttgB* from the evolved isolates had abolished growth completely. These results may be an indication that the isolates were evolved to be more dependent on the mutated copy of *ttgB* to survive on the aromatic acid stressors. This observation agrees with the previous observations that expression this efflux pump is important for tolerance phenotype under stress conditions [40].

There are several identified mutations from this work that need further detailed studies to understand their advantages on the aromatic acid tolerance. In-depth analysis of mutations

identified across the TALE conditions on different substrates, or a mixture of them, can provide insight into the mechanistic impact of specific mutations. It was observed that two genes, *gacS* and *uvrY*, mutated multiple times across different replicates from the different TALE conditions and they are related to the two-component system. Often, the two-component systems enable bacteria to interact and adapt to changes with the surrounding environment or in the intracellular physiology and then induce changes in the transcription [49]. Furthermore, *gacS* (BarA) and *uvrY* (GacA) signal regulates, with other signals, the *csrA* (RsmA) global regulator (carbon storage regulator). This mode of action is preserved across different bacterial species [50] where *csrA*, by itself, regulates numerous unrelated biological pathways in bacterial strains, including oxidative stress (positively) [51], gluconeogenesis (negatively), glycogenesis (negatively) [52], biofilm (negatively) [53] and motility (positively) [54]. Previous studies demonstrated that bacteria under stress by organic solvents adapt by the accumulation of low molecular weight organic solutes such as proline, glycine and trehalose [55,56]. Accumulation of certain carbohydrates in the cytoplasm of bacteria were found to protect the protein and the membrane under stress conditions. For example, *Pseudomonas* strains increased proteins needed for the formation of trehalose in the presence of vanillin [44] and trehalose and glycogen synthesis under osmotic stress [57]. The assumed accumulation of these carbohydrates can serve not only for protection but also as a carbon source under external stress and under certain conditions. The role of each mutation is not clear and requires future work in order to reveal their mechanistic roles of action.

In addition, it was found that several genes related in the tricarboxylic acid cycle (TCA) were mutated across different replicates. In the absence of glucose and presence of aromatic acids as gluconeogenic sources, KT2440 catabolizes *p*-CA and FA via the protocatechuate branch of the β-ketoadipate pathway and TCA intermediates such as acetyl-CoA and succinyl-CoA [13,44,58]. The mutations in the central carbon metabolism genes suggest further activation of the tricarboxylic acid cycle as a response to elevating concentration of aromatic acids towards more energy production and conversion. Previous studies suggested the same response of different *Pseudomonas* strains on exposure to toluene and vanillin [43,44]. All these observations may be efforts to meet the requirement of high energy demand in order to extrude excess aromatic acids or their degradation products. Nevertheless, a

follow up study is still needed to understand the impact of the key mutations individually, or in combination on the tolerance phenotype.

Conditions specific mutations were identified in glucose ALE. Typically, KT2440 cells use glucose as a carbon and energy source where it is metabolized by a cyclic process involving Entner-Doudoroff pathway, parts of the Embden-Meyerhof-Parnas (EMP) pathway and the Pentose phosphate (PP) pathway [59]. Initially, glucose is transported into the periplasm by outer membrane selective porin, OprB, and as expected, several mutations were found in glucose ALE and mixture conditions. Interestingly, *oprB* mutations were observed in mixture TALE even though there was no glucose as a carbon source. This may be further suggesting *oprB*-II to have broad spectrum for transporting carbon sources. Further study is needed to understand the role of *oprB*-II with glucose and in specific with the mixture condition. Another key identified mutation in the control ALE on glucose was in the *gacS* gene, which also has a role in central carbon metabolism. The mechanistic role of each single mutation according to the observed tolerance or fitness improved phenotype is not clear. Additionally, the additive tolerance advantage from having more than one of the key identified mutations requires further follow-up experiments in order to reveal the exact function every mutation has on different growth substrates.

In summary, the utilization of adaptive evolution approach enabled the generation of promising platform strains with higher tolerance level up to 20 and 30 g/L for *p*-CA and *trans*-FA respectively and improved fitness on glucose (up to a growth rate of 0.8 h^{-1}). The approach used to identify and interpret key mutations by analyzing mutations from different populations and clones then identify key mutations in genes mutated several times or the same mutation occurs across different samples was successful in revealing the key mutations behind the tolerance and the fitness improved phenotype. The most remarkable identified key mutations identified in this study appeared to involve in membrane transport mechanisms, possibly exporting the aromatic acids out of the cell, central carbon metabolism, down-regulation of cell motility, and global transcriptional and post-transcriptional regulations. The key mutations identified in this study provide insights into the potential improved tolerance and fitness mechanisms in the evolved strains which can be used readily for engineering tolerance and fitness in a production host strain. The results and generated platform strains and their interesting phenotype can serve as parts for engineering for KT2440 and any other industrially relevant Pseudomonas strains for sustainable production based on raw materials.

List of Supplementary Files

Supplementary Data File 1 – Excel file of all of the whole genome sequencing results. The file can be found here: https://bit.ly/35ralM7

Supplementary Text – Text files contains all supplementary Text, Figures, Tables. The file can be found here: https://bit.ly/2szTubq

Supplementary Data File 2 – Excel file of statistical analysis of the screening results on different substrate conditions. The file can be found here: https://bit.ly/2ssR4v9

Supplementary Data File 3 – Full list of mutations of found mutations in all re-sequenced samples. The file can be found here: https://bit.ly/36LRRGp

Acknowledgements

We would like to thank Patrick Phaneuf, Alexandra Hoffmeyer, and Pannipa Pornpitakpong for their help in preparing the manuscript. We as well want to thank Pablo I. Nikel for providing us with *P. putida* KT2440. Funding for the project was provided by The Novo Nordisk Foundation Grant Number NNF14OC0011269 and Number NNF10CC1016517. The DOE Joint BioEnergy Institute (http://www.jbei.org) is supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U.S. Department of Energy.

References

1. Hoffert MI, Caldeira K, Benford G, Criswell DR, Green C, Herzog H, et al. Advanced technology paths to global climate stability: energy for a greenhouse planet. Science [Internet]. 2002;298:981–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12411695

2. D. Perlack R, Wright L, Turhollow A, Graham R, Stokes B, C. Erbach D. Biomass as Feedstock for A Bioenergy and Bioproducts Industry: The Technical Feasibility of a Billion-Ton Annual Supply. Biomass as Feed. a Bioenergy Bioprod. Ind. Tech. Feasibility a Billion-t. Annu. Supply. 2005.

3. Johnson CW, Beckham GT. Aromatic catabolic pathway selection for optimal production of pyruvate and lactate from lignin. Metab Eng [Internet]. 2015 [cited 2019 Aug 26];28:240–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25617773

4. Zakzeski J, Bruijnincx PCA, Jongerius AL, Weckhuysen BM. The Catalytic Valorization of Lignin for the Production of Renewable Chemicals. Chem Rev [Internet]. American Chemical Society; 2010 [cited 2019 Sep 11];110:3552–99. Available from: https://pubs.acs.org/doi/10.1021/cr900354u

5. Beckham GT, Johnson CW, Karp EM, Salvachúa D, Vardon DR. Opportunities and challenges in biological lignin valorization [Internet]. Curr. Opin. Biotechnol. Elsevier Current Trends; 2016 [cited 2019 Oct 14]. p. 40–53. Available from: https://www.sciencedirect.com/science/article/pii/S0958166916300520

6. Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W. Lignin biosynthesis and structure. Plant Physiol [Internet]. American Society of Plant Biologists; 2010 [cited 2019 Jul 15];153:895–905. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20472751

7. Karp EM, Donohoe BS, O'Brien MH, Ciesielski PN, Mittal A, Biddy MJ, et al. Alkaline Pretreatment of Corn Stover: Bench-Scale Fractionation and Stream Characterization. ACS Sustain Chem Eng [Internet]. American Chemical Society; 2014 [cited 2019 Jul 15];2:1481–91. Available from: http://pubs.acs.org/doi/10.1021/sc500126u

8. Calero P, Jensen SI, Bojanovič K, Lennen RM, Koza A, Nielsen AT. Genome-wide identification of tolerance mechanisms toward p-coumaric acid in Pseudomonas putida. Biotechnol Bioeng [Internet].

2018 [cited 2019 Jun 17];115:762–74. Available from: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bit.26495

9. Salvachúa D, Johnson CW, Singer CA, Rohrer H, Peterson DJ, Black BA, et al. Bioprocess development for muconic acid production from aromatic compounds and lignin. Green Chem [Internet]. The Royal Society of Chemistry; 2018 [cited 2019 Jul 12];20:5007–19. Available from: http://xlink.rsc.org/?DOI=C8GC02519C

10. Bagdasarian M, Lurz R, Rückert B, Franklin FCH, Bagdasarian MM, Frey J, et al. Specific-purpose plasmid cloning vectors II. Broad host range, high copy number, RSF 1010-derived vectors, and a host-vector system for gene cloning in Pseudomonas. Gene [Internet]. Elsevier; 1981 [cited 2019 Jul 14];16:237–47. Available from: https://www.sciencedirect.com/science/article/pii/0378111981900809

11. Baumberg S. Genetics and biochemistry of Pseudomonas. Biochem Educ [Internet]. 2003 [cited 2019 Jul 14];3:46. Available from: https://linkinghub.elsevier.com/retrieve/pii/0307441275900485

12. Clarke PH. The metabolic versatility of pseudomonads. Antonie Van Leeuwenhoek [Internet]. 1982 [cited 2019 Jul 14];48:105–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6808915

13. Jimenez JI, Minambres B, Garcia JL, Diaz E, Jiménez JI, Miñambres B, et al. Genomic analysis of the aromatic catabolic pathways from Pseudomonas putida KT2440. Environ Microbiol [Internet]. John Wiley & Sons, Ltd (10.1111); 2002 [cited 2019 Jun 17];4:824–41. Available from: http://doi.wiley.com/10.1046/j.1462-2920.2002.00370.x

14. Poblete-Castro I, Becker J, Dohnt K, dos Santos VM, Wittmann C. Industrial biotechnology of Pseudomonas putida and related species. Appl Microbiol Biotechnol [Internet]. 2012 [cited 2019 Jul 14];93:2279–90. Available from: http://link.springer.com/10.1007/s00253-012-3928-0

15. Timmis KN, Steffan RJ, Unterman R. Designing Microorganisms for the Treatment of Toxic Wastes. Annu Rev Microbiol [Internet]. 1994 [cited 2019 Jul 14];48:525–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7826017

16. Schmid A, Dordick JS, Hauer B, Kiener A, Wubbolts M, Witholt B. Industrial biocatalysis today and tomorrow. Nature [Internet]. 2001 [cited 2019 Jul 14];409:258–68. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11196655

17. Dejonghe W, Boon N, Seghers D, Top EM, Verstraete W. Bioaugmentation of soils by increasing

microbial richness: missing links. Environ Microbiol [Internet]. 2001 [cited 2019 Jul 14];3:649–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11722545

18. Mohamed ET, Wang S, Lennen RM, Herrgård MJ, Simmons BA, Singer SW, et al. Generation of a platform strain for ionic liquid tolerance using adaptive laboratory evolution. Microb Cell Fact [Internet]. BioMed Central; 2017 [cited 2019 Jun 4];16:1–15. Available from: https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-017-0819-1

19. Mohamed ET, Mundhada H, Landberg J, Cann I, Mackie RI, Nielsen AT, et al. Generation of an E. coli platform strain for improved sucrose utilization using adaptive laboratory evolution. Microb Cell Fact [Internet]. BioMed Central; 2019;18:116. Available from: https://microbialcellfactories.biomedcentral.com/articles/10.1186/s12934-019-1165-2

20. LaCroix RA, Sandberg TE, O'Brien EJ, Utrilla J, Ebrahim A, Guzman GI, et al. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal medium. Appl Environ Microbiol [Internet]. American Society for Microbiology; 2015 [cited 2016 Aug 30];81:17–30. Available from: http://aem.asm.org/cgi/doi/10.1128/AEM.02246-14

21. Sandberg TE, Lloyd CJ, Palsson BO, Feist AM. Laboratory Evolution to Alternating Substrate Environments Yields Distinct Phenotypic and Genetic Adaptive Strategies. Kivisaar M, editor. Appl Environ Microbiol [Internet]. 2017 [cited 2017 Nov 22];83:AEM.00410-17. Available from: http://aem.asm.org/lookup/doi/10.1128/AEM.00410-17

22. Blomfield IC, Vaughn V, Rest RF, Eisenstein BI. Allelic exchange in Escherichia coli using the Bacillus subtilis sacB gene and a temperature-sensitive pSC101 replicon. Mol Microbiol [Internet]. 1991 [cited 2019 Aug 26];5:1447–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1686293

23. Choi K-H, Kumar A, Schweizer HP. A 10-min method for preparation of highly electrocompetent Pseudomonas aeruginosa cells: Application for DNA fragment transfer between chromosomes and plasmid transformation. J Microbiol Methods [Internet]. 2006 [cited 2019 Aug 26];64:391–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15987659

24. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Methods [Internet]. Nature Publishing Group; 2009 [cited 2019 Sep 6];6:343–5. Available from: http://www.nature.com/articles/nmeth.1318

25. Phaneuf PV, Gosting D, Palsson BO, Feist AM. ALEdb 1.0: A Database of Mutations from Adaptive Laboratory Evolution. bioRxiv [Internet]. Cold Spring Harbor Laboratory; 2018 [cited 2019 Jun 4];320747. Available from: https://www.biorxiv.org/content/10.1101/320747v1.full

26. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods [Internet]. Nature Publishing Group; 2012 [cited 2019 Jul 30];9:357–9. Available from: http://www.nature.com/articles/nmeth.1923

27. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol [Internet]. NIH Public Access; 2014 [cited 2019 May 20];1151:165–88. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24838886

28. Winsor GL, Griffiths EJ, Lo R, Dhillon BK, Shay JA, Brinkman FSL. Enhanced annotations and features for comparing thousands of Pseudomonas genomes in the Pseudomonas genome database. Nucleic Acids Res [Internet]. 2016 [cited 2019 Aug 14];44:D646–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26578582

29. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: Towards a more sustainable future. Nucleic Acids Res [Internet]. 2016 [cited 2018 Aug 15];44:D279–85. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26673716

30. Lee DH, Feist AM, Barrett CL, Palsson B. Cumulative number of cell divisions as a meaningful timescale for adaptive laboratory evolution of escherichia coli. Kao KC, editor. PLoS One [Internet]. Public Library of Science; 2011 [cited 2016 Nov 10];6:e26172. Available from: http://dx.plos.org/10.1371/journal.pone.0026172

31. Martínez-Gil M, Ramos-González MI, Espinosa-Urgel M. Roles of cyclic Di-GMP and the Gac system in transcriptional control of the genes coding for the Pseudomonas putida adhesins LapA and LapF. J Bacteriol [Internet]. American Society for Microbiology Journals; 2014 [cited 2019 Jul 2];196:1484–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24488315

32. Saravolac E, Taylor N, Benz R, Hancock R. Purification of glucose-inducible outer membrane protein OprB of Pseudomonas putida and reconstitution of glucose-specific pores. J Bacteriol [Internet]. American Society for Microbiology (ASM); 1991 [cited 2019 Jul 2];173:4970–6. Available from: https://europepmc.org/articles/PMC208185

33. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. Nucleic Acids Res [Internet]. Narnia; 2019 [cited 2019 Oct 15];47:D427–32. Available from: https://academic.oup.com/nar/article/47/D1/D427/5144153

34. Hancock REW, Brinkman FSL. FUNCTION OF PSEUDOMONAS PORINS IN UPTAKE AND EFFLUX. Annu Rev Microbiol [Internet]. 2002 [cited 2019 Aug 9];56:17–38. Available from: http://tigrblast.tigr.org/ufmg/index.cgi

35. Hay ID, Rehman ZU, Rehm BHA. Membrane topology of outer membrane protein AlgE, which is required for alginate production in Pseudomonas aeruginosa. Appl Environ Microbiol [Internet]. American Society for Microbiology; 2010 [cited 2019 Aug 9];76:1806–12. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20097812

36. Bakkevig K, Sletta H, Gimmestad M, Aune R, Ertesvåg H, Degnes K, et al. Role of the Pseudomonas fluorescens Alginate Lyase (AlgL) in Clearing the Periplasm of Alginates Not Exported to the Extracellular Environment. J Bacteriol [Internet]. American Society for Microbiology (ASM); 2005 [cited 2019 Aug 9];187:8375. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16321942

37. Rand JM, Pisithkul T, Clark RL, Thiede JM, Mehrer CR, Agnew DE, et al. A metabolic pathway for catabolizing levulinic acid in bacteria. Nat Microbiol [Internet]. NIH Public Access; 2017 [cited 2019 Jul 31];2:1624–34. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28947739

38. Blanco-Romero E, Redondo-Nieto M, Martínez-Granero F, Garrido-Sanz D, Ramos-González MI, Martín M, et al. Genome-wide analysis of the FleQ direct regulon in Pseudomonas fluorescens F113 and Pseudomonas putida KT2440. Sci Rep [Internet]. Nature Publishing Group; 2018 [cited 2019 Jul 2];8:13145. Available from: http://www.nature.com/articles/s41598-018-31371-z

39. Corral-Lugo A, De la Torre J, Matilla MA, Fernández M, Morel B, Espinosa-Urgel M, et al. Assessment of the contribution of chemoreceptor-based signalling to biofilm formation. Environ Microbiol [Internet]. John Wiley & Sons, Ltd (10.1111); 2016 [cited 2019 Jul 15];18:3355–72. Available from: http://doi.wiley.com/10.1111/1462-2920.13170

40. Basler G, Thompson M, Tullman-Ercek D, Keasling J. A Pseudomonas putida efflux pump acts on short-chain alcohols. Biotechnol Biofuels [Internet]. BioMed Central; 2018 [cited 2019 Aug 6];11:136. Available from: https://biotechnologyforbiofuels.biomedcentral.com/articles/10.1186/s13068-018-1133-9

41. Ramos JL, Duque E, Godoy P, Segura A. Efflux pumps involved in toluene tolerance in Pseudomonas putida DOT-T1E. J Bacteriol [Internet]. American Society for Microbiology (ASM); 1998 [cited 2019 Aug 8];180:3323–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9642183

42. Junker F, Rodríguez-Herva JJ, Duque E, Ramos-González MI, Llamas M, Ramos JL. A WbpL mutant of Pseudomonas putida DOT-T1E strain, which lacks the O-antigenic side chain of lipopolysaccharides, is tolerant to organic solvent shocks. Extremophiles [Internet]. 2001 [cited 2019 Jul 15];5:93–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11354460

43. Wijte D, Van Baar BLM, Heck AJR, Maarten Altelaar AF, Altelaar AFM. Probing the Proteome Response to Toluene Exposure in the Solvent Tolerant Pseudomonas putida S12. J Proteome Res [Internet]. American Chemical Society; 2011 [cited 2019 Aug 5];10:394–403. Available from: http://pubs.acs.org/doi/abs/10.1021/pr100401n

44. Simon O, Klaiber I, Huber A, Pfannstiel J. Comprehensive proteome analysis of the response of Pseudomonas putida KT2440 to the flavor compound vanillin. J Proteomics [Internet]. 2014 [cited 2019 Jun 17];109:212–27. Available from: http://dx.doi.org/10.1016/j.jprot.2014.07.006

45. Ito F, Tamiya T, Ohtsu I, Fujimura M, Fukumori F. Genetic and phenotypic characterization of the heat shock response in Pseudomonas putida. Microbiologyopen [Internet]. Wiley-Blackwell; 2014 [cited 2019 Jul 15];3:922–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25303383

46. Volkers RJM, Snoek LB, Ruijssenaars HJ, de Winde JH. Dynamic Response of Pseudomonas putida S12 to Sudden Addition of Toluene and the Potential Role of the Solvent Tolerance Gene trgI. PLoS One [Internet]. Public Library of Science; 2015 [cited 2019 Jul 15];10:e0132416. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26181384

47. Cuenca M del S, Roca A, Molina-Santiago C, Duque E, Armengaud J, Gómez-Garcia MR, et al. Understanding butanol tolerance and assimilation in Pseudomonas putida BIRD-1: an integrated omics approach. Microb Biotechnol [Internet]. Wiley-Blackwell; 2016 [cited 2019 Jul 15];9:100. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26986205

48. Hishinuma S, Ohtsu I, Fujimura M, Fukumori F. OxyR is involved in the expression of thioredoxin reductase TrxB in Pseudomonas putida. FEMS Microbiol Lett [Internet]. Narnia; 2008 [cited 2019 Oct 16];289:138–45. Available from: https://academic.oup.com/femsle/article-lookup/doi/10.1111/j.1574-6968.2008.01374.x

49. Timmermans J, Van Melderen L. Post-transcriptional global regulation by CsrA in bacteria. Cell Mol Life Sci [Internet]. SP Birkhäuser Verlag Basel; 2010 [cited 2019 Aug 7];67:2897–908. Available from: http://link.springer.com/10.1007/s00018-010-0381-z

50. Heeb S, Blumer C, Haas D. Regulatory RNA as Mediator in GacA/RsmA-Dependent Global Control of Exoproduct Formation in Pseudomonas fluorescens CHA0. J Bacteriol [Internet]. 2002 [cited 2019 Aug 8];184:1046–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11807065

51. Carmel-Harel O, Storz G. Roles of the Glutathione- and Thioredoxin-Dependent Reduction Systems in the Escherichia Coli and Saccharomyces Cerevisiae Responses to Oxidative Stress. Annu Rev Microbiol [Internet]. 2000 [cited 2019 Aug 7];54:439–61. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11018134

52. Romeo T, Gong M, Mu Ya Liu, Brun-Zinkernagel AM. Identification and molecular characterization of csrA, a pleiotropic gene from Escherichia coli that affects glycogen biosynthesis, gluconeogenesis, cell size, and surface properties. J Bacteriol [Internet]. 1993 [cited 2019 Aug 7];175:4744–55. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8393005

53. Jackson DW, Suzuki K, Oakford L, Simecka JW, Hart ME, Romeo T. Biofilm formation and dispersal under the influence of the global regulator CsrA of Escherichia coli. J Bacteriol [Internet]. 2002 [cited 2019 Aug 7];184:290–301. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11741870

54. Wei BL, Brun-Zinkernagel A-M, Simecka JW, Prüß BM, Babitzke P, Romeo T. Positive regulation of motility and flhDC expression by the RNA-binding protein CsrA of Escherichia coli. Mol Microbiol [Internet]. 2001 [cited 2019 Aug 7];40:245–56. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11298291

55. Isken S, de Bont JA. Bacteria tolerant to organic solvents. Extremophiles [Internet]. 1998 [cited 2019 Aug 6];2:229–38. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9783170

56. Kempf B, Bremer E. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. Arch Microbiol [Internet]. 1998 [cited 2019 Aug 6];170:319–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9818351

57. Freeman BC, Chen C, Beattie GA. Identification of the trehalose biosynthetic loci of Pseudomonas syringae and their contribution to fitness in the phyllosphere. Environ Microbiol [Internet]. 2010 [cited

2019 Aug 6];12:1486–97. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20192963

58. Kim YH, Cho K, Yun SH, Kim JY, Kwon KH, Yoo JS, et al. Analysis of aromatic catabolic pathways in Pseudomonas putida KT 2440 using a combined proteomic approach: 2-DE/MS and cleavable isotope-coded affinity tag analysis. Proteomics [Internet]. 2006 [cited 2019 Aug 6];6:1301–18. Available from: www.proteomics-journal.com

59. Nikel PI, Chavarría M, Fuhrer T, Sauer U, De Lorenzo V. Pseudomonas putida KT2440 strain metabolizes glucose through a cycle formed by enzymes of the Entner-Doudoroff, embden-meyerhof-parnas, and pentose phosphate pathways. J Biol Chem. American Society for Biochemistry and Molecular Biology Inc.; 2015;290:25920–32.