Aalto University
School of Chemical
Engineering

Tuula Tenkanen

METABOLIC MODEL GUIDED DESIGN ENABLING ADAPTIVELY EVOLVING IMPROVED PRODUCTION BY MICROBIAL HOSTS

Master´s Programme in Chemical, Biochemical and Materials Engineering
Major in Biotechnology

Master’s thesis for the degree of Master of Science in Technology submitted for inspection, Espoo, 24 May, 2021.

Supervisor	Professor Merja Penttilä
Instructor	D.Sc. (Tech.) Paula Jouhten

Aalto University
School of Chemical Engineering

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi

Abstract of master's thesis

Author Tuula Tenkanen

Title of thesis Metabolic model guided design enabling adaptively evolving improved production by microbial hosts

Degree Programme Master's Programme in Chemical, Biochemical and Materials engineering

Major Biotechnology

Thesis supervisor Professor Merja Penttilä

Thesis advisor(s) / Thesis examiner(s) D.Sc. (Tech.) Paula Jouhten

Date 24.05.2021 **Number of pages** 63+17 **Language** English

Abstract

Engineering microbial hosts to achieve higher production levels of e.g. chemicals is challenging as genotype-phenotype relationships are limitedly known. Adaptive laboratory evolution allows host improvement without the need to know the genotypic underpinnings of phenotypes. However, adaptive laboratory evolution can only be used when the desired production related trait is genetically correlated with fitness (i.e. growth), and chemical production is rarely naturally growth-coupled.

Metabolic network creates couplings between fluxes, including biosynthetic fluxes for growth. Besides the metabolic network structure, also the nutritional environment affects the flux couplings, and these can be predicted using genome-scale metabolic model simulations. In this work, a novel algorithm EvolveXGA that use genome-scale metabolic models was demonstrated for designing strategies combining biochemical reaction deletions (i.e. achievable with gene deletions) and nutritional environments to optimize target flux(es) coupling with growth. When the target fluxes are chosen to include either the production pathway or reactions providing precursors for the desired pathway, the strategies are predicted to enable adaptively evolving improved production.

EvolveXGA uses a genetic algorithm as an optimization routine. First, the genetic algorithm was set up for optimizing the target flux(es) coupling to growth using reaction deletions and nutritional environment as variables. Then, EvolveXGA was tested for designing strategies including reaction deletions and nutritional environment to growth-couple glycolic acid production pathway in *Saccharomyces cerevisiae*. Finally, the generalizability of the method was evaluated by designing such strategies for 28 other heterologous products. EvolveXGA found strategies to couple 13 of these production pathways with growth including the oxalate pathway for glycolic acid. In addition, for the remainder of the pathways EvolveXGA found strategies to couple native precursor providing reactions with growth. Some of these strategies showed potential to be realized *in vivo*.

Synthetic biology tools have enabled fast development and implementation of new production pathways into microbial hosts. However, improving the production experimentally from lab demonstration to economically feasible levels has remained resource-intensive and creates a bottleneck for new industrial microbial fermentation processes. The novel model-guided approach that was demonstrated *in silico* in this project offers generalizable means to accelerate the production improvement using operationally simple, powerful, adaptive laboratory evolution.

Keywords strain improvement, metabolic modelling, adaptive laboratory evolution, EvolveX algorithm, EvolveXGA algorithm, genetic algorithm

Aalto-yliopisto
Kemian tekniikan korkeakoulu

Aalto-yliopisto, PL 11000, 00076 AALTO
www.aalto.fi

Diplomityön tiivistelmä

Tekijä Tuula Tenkanen

Työn nimi Laboratorioevoluutiostrategioiden suunnittelu aineenvaihduntamallien avulla mikrobikantojen tuoton parantamisen mahdollistamiseksi

Koulutusohjelma Master's Programme in Chemical, Biochemical and Materials engineering

Pääaine Biotechnology

Työn valvoja Professori Merja Penttilä

Työn ohjaaja(t)/Työn tarkastaja(t) TkT Paula Jouhten

Päivämäärä 24.5.2021 **Sivumäärä** 63+17 **Kieli** Englanti

Tiivistelmä

Tuotantokantojen kehitys parempien tuotantotasojen saavuttamiseksi on hankalaa, koska genotyypin ja fenotyyppin välistä riippuvuutta ei usein tunneta riittävän hyvin. Adaptiivinen laboratorioevoluutio mahdollistaa kantojen kehityksen ilman, että genotyypin ja fenotyypin välistä yhteyttä tarvitsisi tietää. Adaptiivista laboratorioevoluutiota voi kuitenkin käyttää ainoastaan silloin, kun haluttu ominaisuus on geneettisesti kytköksissä kasvuun, mutta haluttu tuotanto on harvoin suoraan kasvuun kytkeytynyt.

Aineenvaihduntaverkko muodostaa kytköksiä aineenvaihdunnan reaktioiden välille. Aineenvaihduntaverkon lisäksi myös kasvuympäristö vaikuttaa siihen, mitkä reaktiot ovat kytköksissä. Reaktioiden kytköksiä on mahdollista ennustaa genominlaajuisia aineenvaihduntamalleja simuloimalla. Tässä työssä esitellään uusi algoritmi, EvolveXGA, joka hyödyntää genominlaajuisia aineenvaihduntamalleja ja suunnittelee strategioita, joilla halutut tavoitereaktiot voidaan kytkeä solujen kasvuun. Nämä strategiat sisältävät sekä reaktiodeleetioita, jotka voidaan toteuttaa geenideleetioilla, että sopivan kasvuympäristön määrittelyn. Oletuksena on, että EvolveXGAn määrittelemät strategiat mahdollistavat tuotannon parantumisen adaptiivisen laboratorioevoluution aikana, kun tavoite reaktiot valitaan siten, että ne ovat osa tuotteeseen johtavaa aineenvaihduntareittiä tai tuottavat lähtöaineita tuotantoreitille.

EvolveXGA käyttää optimoinnissa geneettistä algoritmia. Ensin geneettinen algoritmi alustettiin tavoitereaktioiden kasvuun kytkennän optimoimiseksi, reaktiodeleetiot ja kasvuympäristö muuttujina. Tämän jälkeen kokeiltiin, pystyykö EvolveXGA löytämään reaktiodeleetioita ja kasvuympäristön sisältäviä strategioita siten, että glykolihapon tuottoreitti kytkeytyy kasvuun *Saccharomyces cerevisiae* hiivassa. Lopuksi, lähestymistavan yleistettävyyttä evaluoitiin suunnittelemalla näitä strategioita 28 muulle heterologiselle tuotteelle. EvolveXGA löysi strategiat kytkeä 13 heterologisen tuotteen, sisältäen glykolihapon, synteesireitit kasvuun. Lisäksi niille tuotteille, joille synteesireitin kytkeminen kasvuun ei onnistunut, EvolveXGA löysi tavan kytkeä hiivan omia, synteesireitin lähtöaineita tuottavia reaktioita kasvuun. Osa EvolveXGAn löytämistä strategioista voisivat myös olla toteutettavissa.

Synteettinen biologia on mahdollistanut uusien tuotteiden synteesireittien nopean kehityksen ja viennin mikrobeihin laboratoriossa. Tuottotasot ovat usein kuitenkin liian matalia, jotta teollisen mittakaavan tuotanto olisi kannattavaa. Tässä työssä esitellyn uuden, aineenvaihduntamallejamalleja hyödyntävän, lähestymistavan odotetaan mahdollistavan tuottotasojen parantamisen helposti toteutettavalla adaptiivisella laboratorioevoluutiolla.

Avainsanat tuotantokantojen kehitys, metabolian mallinnus, adaptiivinen laboratorioevoluutio, EvolveX algoritmi, EvolveXGA algoritmi, geneettinen algoritmi

Preface

This master's thesis project was performed between May 2020 and May 2021 in the group of Bioanalytics and biological data science at the Technical Research Centre of Finland Ltd (VTT Ltd). The aim of the work was to demonstrate a computational algorithm (EvolveXGA) for searching reaction deletions and selection niche to achieve growth-coupling of selected target reactions. EvolveXGA had been developed by Paula Jouhten and Sandra Castillo had performed the initial python implementation of EvolveXGA.

First, I want to thank my supervisor Merja Penttilä and my advisor Paula Jouhten for offering me the possibility to work with this interesting topic. In addition, I would like to thank Merja Penttilä for her support and feedback during the thesis process and Paula Jouhten for her continuous support and will to share her knowledge. I learnt more than I could have imagined during this process. I would also want to thank Sandra Castillo, especially for her help with python language and the algorithm implementation.

Finally, I would like to thank my family for always supporting me, Ville for patiently following the process and teaching some excel skills and Jon, Jocke, Mikke and everyone else in Sailing Team Finland for their understanding and support.

Espoo 2021

Tuula Tenkanen

Table of contents

List of symbols and abbreviations

$\boldsymbol{a}$	vector of possible nutrient uptakes
$\boldsymbol{c}$	vector of weights representing how each flux contributes to the objective function
D	set of metabolic reactions
$e(t)$	enzyme concentration
p	kinetic parameters
$r_{uptake,max}$	optimal nutrition uptake
$\boldsymbol{S}$	stoichiometric matrix
$s(t)$	reactant concentration
T	set of upregulation targets
U	set of absolute flux variables of the upregulation targets
$\boldsymbol{v}$	vector of fluxes
$\boldsymbol{v_{del}}$	vector of deleted reaction fluxes
v_{μ}	flux of biomass production
$\boldsymbol{v_{lb}}$	vector of flux lower bounds
$\boldsymbol{v_{ub}}$	vector of flux upper bounds
$v_{\|u\|,ub}$	upper bound of the absolute flux variable of a upregulation target |u|
$w_{lb,i}$	lower threshold flux i of wildtype
$w_{\mathrm{max,i}}$	maximum flux i in wildtype
$w_{\mathrm{min,i}}$	minimum flux i in wildtype
$w_{\mathrm{ub,i}}$	upper threshold flux i of wildtype
$\boldsymbol{X}$	vector of metabolite concentrations
y_i	binary variable defining whether a reaction flux i has to be up- or downregulated for the production to reach a level α of maximum production

ALE	Adaptive laboratory evolution
AMP	Adenosine monophosphate
ATP	Adenosine triphosphate
CDW	Cell dry weight
FBA	Flux balance analysis
FVA	Flux variability analysis
GEM	Genome-scale metabolic model
LP	Linear programming
MILP	Mixed-integer linear programming
MoMA	Minimization of Metabolic Adjustment
NAD^+	Nicotinamide adenine dinucleotide
NADH	Reduced form of NAD^+
$NADP^+$	Nicotinamide adenine dinucleotide phosphate
NADPH	Reduced form of $NADP^+$
pFBA	Parsimonious flux balance analysis
ROOM	Regulatory On/Off Minimization
TCA cycle	Tricarboxylic acid cycle

1 Introduction

Climate change and depletion of fossil raw materials create a continuously increasing need to develop environmentally friendly means for producing chemical compounds. Manufacturing of compounds with microbial fermentation enable the use of renewable raw materials and milder production conditions compared to chemical processes. High production levels by the microbial cells are needed for economical feasibility and attractiveness. Therefore, strain development to optimize production levels is essential.

Strain improvement is often a slow trial and error process as it is scarcely known how the genotype should be modified to achieve a desired phenotype (e.g. high product yield or efficient use of a certain substrate). Adaptive laboratory evolution (ALE) is a powerful tool in strain improvement since it circumvents the need of knowing the genetic underpinnings behind improved phenotypes. In ALE cells are cultured for a prolonged time and fitness increasing mutations become enriched (Dragosits and Mattanovich, 2013; Sandberg *et al.*, 2019). Thus, a limitation with ALE is that it can be used to improve the wanted traits only when they are coupled with fitness (i.e. growth) and production pathways, heterologous in particular, are rarely growth-coupled as such.

One way to achieve growth-product coupling is to find suitable metabolic network modifications (i.e. gene deletions or flux up- or downregulation) in a way that the production becomes an obligatory byproduct of (optimal) growth. Several algorithms that aim to define these modification targets have been developed (Bugard *et al.*, 2003; Pharkya and Maranas, 2006; Tepper and Shlomi, 2010; Yang e*t al.*, 2011). Yet, growth-product coupling with network modifications has not been possible for all compounds (Campodonico *et al.*, 2014; Feist *et al.*, 2010). However, in addition to the metabolic network structure, also nutritional environment affects how fluxes are distributed in the metabolic network for optimal cellular fitness (Pál and Papp, 2017; Szappanos *et al.*, 2016).

Recently, a novel approach using the nutrition-dependence of the flux couplings was developed for adaptively evolving desired traits, that are not necessarily fitness increasing in the *target niche* (i.e. chemical environment in the target application e.g. production of a desired compound) (Jouhten and Patil, 2016; Jouhten *et al*., 2021 Preprint). The approach involves a novel algorithm (EvolveX) simulating a genome-scale metabolic model to design a *selection niche* (i.e. chemical environment for ALE) that couples selected target fluxes (e.g. reactions of a production pathway) with fitness (i.e. growth) through the metabolic network.

The aim of this work is to test and evaluate *in silico* an algorithmic extension of EvolveX (EvolveXGA) that combines the design of a selection niche and metabolic reaction deletions (i.e. achievable with gene deletions) to create a coupling between selected target fluxes and fitness. This novel combined strategy is expected to broaden the coverage of fluxes and pathways that can be exposed to selection in ALE.

EvolveXGA aims to optimize the growth-coupling arising from the selection niche and reaction deletions and is implemented as a genetic algorithm with EvolveX as the fitness function. Genetic algorithms are general metaheuristic optimization algorithms that mimic an evolutionary optimization process. The nomenclature of the objects and processes in the genetic algorithm are also adopted from biological evolution. The genetic algorithm term for a candidate solution is a “chromosome” of an “individual”. A pool of such candidate solutions is called a “population”. First, an initial population is created by forming individuals with chromosomes (i.e. different candidate solutions). The fitness of each individual is then scored with a fitness function that suits the particular optimization task. In EvolveXGA the EvolveX scoring algorithm is used as a fitness function. After the fitness of each candidate solution is evaluated, a new population is formed through “elitism”, “crossovers” and “mutations”. Elitism process transfers a defined proportion of best candidate solutions to the new population. Then the rest of the new population is generated through crossover and mutation. In crossover new combinations of current candidate solutions are formed and mutations generate changes to the offspring solutions originating from crossing over. After forming the new population the individuals are again scored and this loop of creating new solutions and scoring them is continued until a predetermined number of generations or fitness is reached. Therefore, when using a genetic algorithm it is not necessary to rank the whole search space to find an optimal solution. The fitness improvement in EvolveXGA would indicate achieving a design that exposes the target fluxes to stronger selection in ALE.

EvolveXGA was first evaluated against an intuitive reaction deletion strategy designed by Paula Jouhten for improving glycolic acid production in *Saccharomyces cerevisiae* using ALE. After this, the generalizability of the method was validated by searching combined selection niche and reaction deletion strategies for 29 heterologous production pathways, including the production pathway for glycolic acid, drawing their main precursors from different parts of central metabolism. Each of these heterologous production pathways was individually introduced to the genome-scale metabolic model of *S. cerevisiae* and these models were used for EvolveXGA optimizations. The primary hypothesis was that EvolveXGA can find a suitable selection niche and reaction deletions to couple selected target reactions with fitness and therefore is predicted to enable flux improvement of these selected target reactions during ALE. The secondary hypothesis was that these growth-coupling designs can be found for all the 29 heterologous production pathways that were selected to be tested.

This thesis starts with a background part that focuses on metabolic model guided strain improvement. After this the used research methods and attained results are presented. The applicability of the attained growth-coupling solutions for *in vivo* strain improvement is discussed and finally conclusions about the results are presented.

2 Background

2.1 Microbes as production hosts

Microbes are already used for industrial production of various compounds like antibiotics (Najafpour, 2015), enzymes (Niyonzima, 2019), and biofuels (Vohra *et al.*, 2014). Yet, huge underused potential remains in microbial metabolism for production of a wide variety of other industrially relevant compounds (Lee *et al.*, 2019; Zhang *et al.*, 2016). Production possibilities are not limited to the native production pathways. Synthetic biology tools (i.e. genome editing and gene synthesis) enable assembling completely novel pathways from enzymes with different origins. The novel pathways can be designed by combining existing reactions from different species or even by designing novel enzyme mechanisms allowing the development of microbial fermentation processes with new-to-nature reactions (Erb *et al.*, 2017).

Microbial fermentation processes have several advantages over conventional chemical production largely relying on petrochemical fractions. An important benefit is that microbes can use renewable raw materials for chemical synthesis. The renewable feedstocks for microbial chemical synthesis include readily available sugars like glucose, fructose, sucrose and xylose; lignocellulosic raw materials; oils and municipal solid wastes (Manzer e*t al.*, 2013), or even CO_2 (Zabranska and Pokorna, 2018). The ones not competing with food production (e.g. lignocellulosic raw material, waste streams and CO_2) are preferable. Another benefit of microbial fermentation processes is that, compared to chemical processes, fermentations consume less harmful chemical reagents, in general need smaller number of unit operations, are highly regio- and stereoselective, and are safer to operate as the reaction conditions are milder (e.g. ambient temperature and pressure) (Lee *et al.*, 2019).

However, achieving industrially feasible production levels (i.e. yield, productivity, titer) is often a challenge in the microbial chemical synthesis. Therefore, industrial scale production has not realized for many compounds for which the production routes have been established in laboratory (Wehrs *et al.*, 2019). In nature, the production of byproducts in high yields is usually not beneficial. In addition, when a heterologous production pathway is used, the native metabolism might not direct sufficient resources to the desired production pathway for reaching feasible production levels. Strain improvement is thus needed to direct resources from available substrates better to the desired product.

Strain improvement by intuitive design carried out by a scientist is difficult as the genetic underpinnings behind phenotypic traits are often unknown. Genome-scale metabolic models (GEMs) allow simulations of the whole metabolism (Figure 1A) and therefore, they can be used to design redirection of cellular resources to the wanted product.

2.2 Genome-scale metabolic model-guided host improvement

2.2.1 Genome-scale metabolic models

Genome-scale metabolic models (GEMs) have been constructed for more than 6000 organisms (Gu *et al.*, 2019). Manually curated models have been developed for model organisms such as *Escherichia coli* and *S. cerevisiae* as well as for other industrially important microorganisms (e.g. *Corynabacterium glutamicum*, *Pichia pastoris*). In addition, GEMs can be automatically reconstructed from genome data (Gu *et al.*, 2019).

GEMs aim at describing all the biochemical conversions that the organism is capable of performing as a mathematical representation. In this mathematical representation reaction stoichiometries are presented as a stoichiometric matrix (Figure 1B). The metabolic capabilities are encoded in the organism's genome since genes encode metabolic enzymes that catalyze the biochemical conversions (i.e. gene-protein-reaction rule).

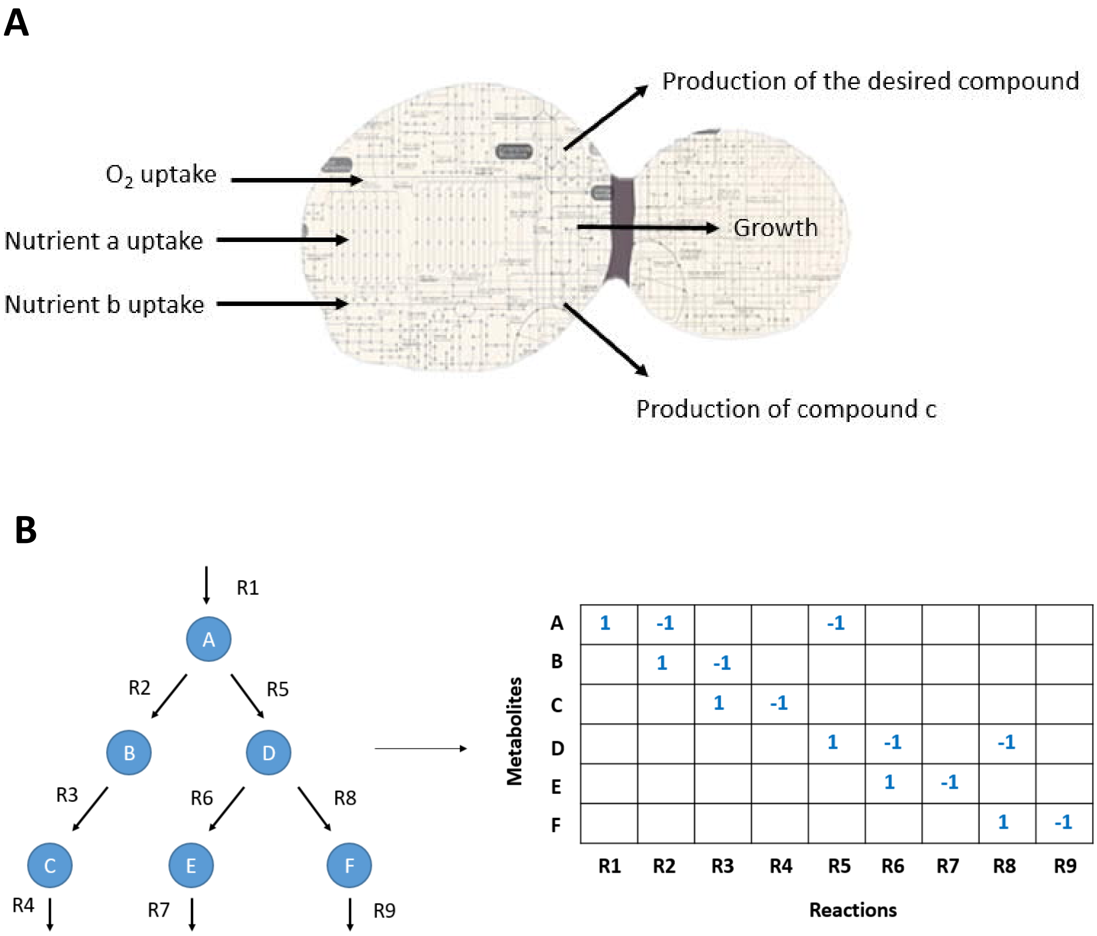

Figure 1. A) GEMs allow simulations of the whole metabolism. Cell figure template obtained from Paula Jouhten. B) Stoichiometries of metabolic reactions are presented as a stoichiometric matrix. A-F represents metabolites and R1-R9 metabolic reactions. Figure modified from O'Brien *et al*. (2015).

The reaction stoichiometries impose mass conservation constraints on the metabolism. Dynamic mass balance equations for metabolites can be formulated as the differences between the metabolite consumption and production rates (Edwards *et al*., 2002; Varma and Palsson, 1994). When the mass balance equations for all the metabolites in a cell are considered a matrix representation is attained (Equation 1).

$$\frac{dX}{dt} = \boldsymbol{S} \cdot \boldsymbol{v} = \boldsymbol{S} \cdot \boldsymbol{f}(\boldsymbol{e}(\boldsymbol{t}), \boldsymbol{s}(\boldsymbol{t}), \boldsymbol{p}) \qquad \text{(Equation 1)}$$

where ***X*** is a vector of metabolite concentrations

S is a matrix containing the stoichiometries of the metabolic reactions

v is a vector representing the fluxes through all the reactions in the network

e(t) is enzyme concentration

s(t) is reactant concentration

p is enzyme parameters

The kinetic mechanisms (i.e. the functions *f(e(t),s(t),p)* and parameter values (*p*)) to define fluxes of many enzymatic reactions are nevertheless not known. However, metabolic transients (kinetics) are much faster than cellular growth or environmental changes and therefore a steady state assumption is often well justified. The steady state assumption is particularly justified for a steady state growth in a chemostat (i.e. under constant conditions and chemical environment) and in the exponential phase of a batch culture since environmental conditions do not restrict growth. The steady state assumption leads to a linear flux balance equation (Equation 2) (Edwards e*t al.*, 2002).

$$\boldsymbol{S} \cdot \boldsymbol{v} = 0 \qquad \text{(Equation 2)}$$

Further, ultimate physicochemical constraints on metabolic fluxes due to thermodynamics can be derived from Gibbs free energies of reactions and introduced into the GEMs as reaction upper and lower bounds defining reaction directionality (Castillo *et al*., 2019; Edwards *et al*., 2002). The thermodynamic constraints or any known enzyme capacity constraints together with the mass conservation constraints and the steady state assumption form the ultimate limits of feasible metabolic phenotypes (Figure 2).

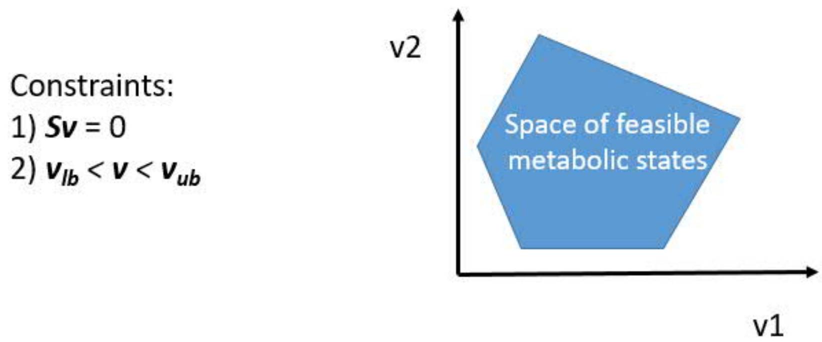

Figure 2. Feasible metabolic phenotypes. The mass balance equation and reaction flux bounds form the constraints of possible fluxes for each reaction. For simplicity only two reactions are drawn to the figure, but in reality the allowable solution space is formed with the fluxes of all the metabolic reactions in the GEM. $\boldsymbol{v_{lb}}$ and $\boldsymbol{v_{ub}}$ represents flux lower and upper bounds respectively. Figure modified from O'Brien *et al*. (2015).

2.2.2 Metabolic phenotype prediction methods

The mathematical formulation of the space of feasible metabolic states arising from the particular metabolic genotype allows predictions of likely and possible metabolic phenotypes. Single metabolic phenotypes (i.e. flux states) are points in the allowable space and can be explored by sampling or by identifying points of optimality. Points of optimality are located in the corners of the space and can be identified using linear optimization as due to the steady state assumption, all the constraints in the GEMs rendered linear. Flux balance analysis (FBA) formulates the simplest form of a linear optimization (Equation 3) (Varma and Palsson, 1994; Orth e*t al*., 2010) for identifying optimal metabolic phenotypes using GEMs. The objective function for FBA is chosen according to the kind of phenotype to be identified (e.g. optimal growth or maximum theoretical rate/yield) and the optimal solution is the solution in the allowable solution space that maximizes the objective function (Figure 3).

maximize (or minimize) $\boldsymbol{c} \cdot \boldsymbol{v}$ (Equation 3)

s.t. $\boldsymbol{S} \cdot \boldsymbol{v} = 0$

$\boldsymbol{v}_{lb} < \boldsymbol{v} < \boldsymbol{v}_{ub}$

where $\boldsymbol{c}$ is a vector of weights representing how each flux contributes to the objective function.

$\boldsymbol{v}_{lb}$ and $\boldsymbol{v}_{ub}$ are flux lower and upper bounds, respectively

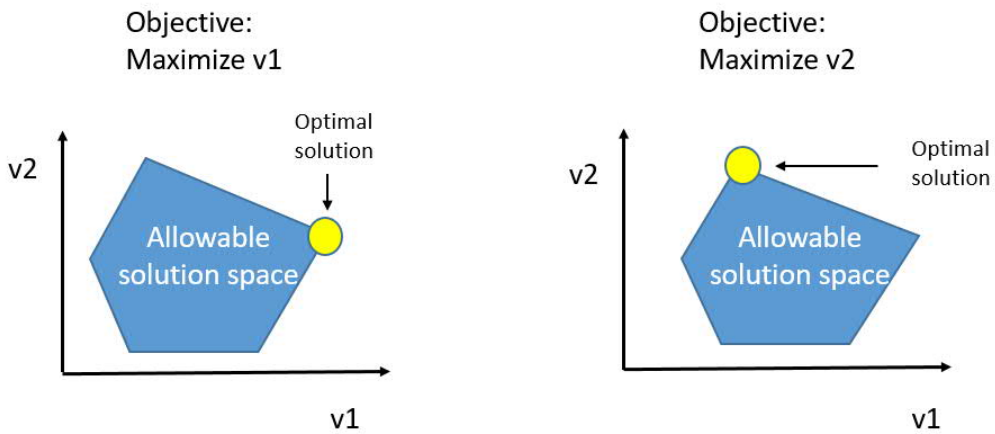

Figure 3. FBA solutions. The result of an FBA is a flux distribution that satisfies the flux constraints and maximizes (or minimizes) the objective function. The objective function can consist of one flux (as in the figure) or of several fluxes. For simplicity only two reactions are drawn to the figure, but in reality the allowable solution space is formed with the fluxes of all the metabolic reactions in the GEM. Figure modified from O'Brien *et al*. (2015).

Maximum theoretical rate/yield can be determined by maximizing the product formation (Varma *et al*., 1993). Optimal rates can be predicted if true nutrient uptake rates are available. Otherwise, arbitrary uptake fluxes can be used which allows the estimation of optimal yields (on the limiting substrate). For predicting native metabolic states of microbes, using a maximization of growth as an objective function has been found to well reproduce experimental data (Edwards *et al*., 2001; Ibarra *et* al., 2002; Segrè *et al.*, 2002). Other objective functions have also been suggested for simulating true biological states like maximizing ATP generation (Kauffman *et al.*, 2003).

Growth maximization simulates well the native microbial metabolic phenotypes as they are optimized through evolution in nature (i.e. in microbial populations mutations that increase growth rate naturally enrich and growth becomes maximized). However, this assumption might not be reasonable for mutants generated in the laboratory as they have not been under evolutionary pressure similarly as the wild types (Segrè *et al.*, 2002). Therefore, Segrè *et al.* (2002) suggested another simulation algorithm, Minimization of Metabolic Adjustment (MoMA), for predicting metabolic phenotypes of mutant strains after gene knockouts. In MoMA it is assumed that the regulation of metabolism prevails and thus the flux distribution of a mutant strain remains as close as possible to the wild type flux distribution. Therefore, the mutant metabolic phenotypes are predicted by solving a quadratic optimization problem minimizing the distance in the flux space between the wild type fluxes and the mutant fluxes. An FBA solution or an experimentally defined wild-type flux state can be used as the wild-type fluxes (Segrè *et al.*, 2002).

An alternative algorithm, Regulatory On/Off Minimization (ROOM) for predicting mutant metabolic phenotypes was presented by Shlomi *et al.* (2005). ROOM uses mixed integer linear programming to minimize the number of metabolic reactions whose fluxes need to substantially change compared to the wild type flux state. Thus the underlying assumption in ROOM is that a cell minimizes the genetic regulatory changes that are needed to increase or decrease fluxes and in that way minimizes the cost of adaption (Shlomi *et al*., 2005).

However, the flux state given by these methods is just one possible flux state that optimizes the particular objective function. The optimal objective function value is globally optimal but might be reached also with other flux states. Parsimonious flux balance analysis (pFBA) presented by Lewis *et al*. (2010) aims at finding the most biologically relevant flux state by first defining the optimal growth rate/yield with FBA and then minimizing the sum of the fluxes of all reactions while keeping the growth rate optimal. The assumption behind pFBA is that the total enzyme mass is minimized in evolutionary optimized cells and this can be approximated with minimizing the flux through the metabolic network (Lewis *et al.*, 2010). However, there might be several optimal solutions even with pFBA (Castillo *et al.*, 2019).

With flux variability analysis (FVA) the possible flux variation for each reaction satisfying the objective value can be determined (Bugard *et al.*, 2001; Mahadevan and Schilling, 2003). FVA determines minimum and maximum values of each flux given a particular objective function value. Therefore, FVA can be used e.g. to check if a flux can vary in optimal FBA solutions (Maia *et al.*, 2016), or to see which reactions are necessary e.g. for growth or a desired product formation.

As genotype-phenotype relationships can be predicted using the genome-scale metabolic model simulations with the discussed phenotype prediction methods, the methods can be used for designing strain improvement strategies. Such strategies may involve for example gene deletions, to achieve a desired phenotype (Figure 4).

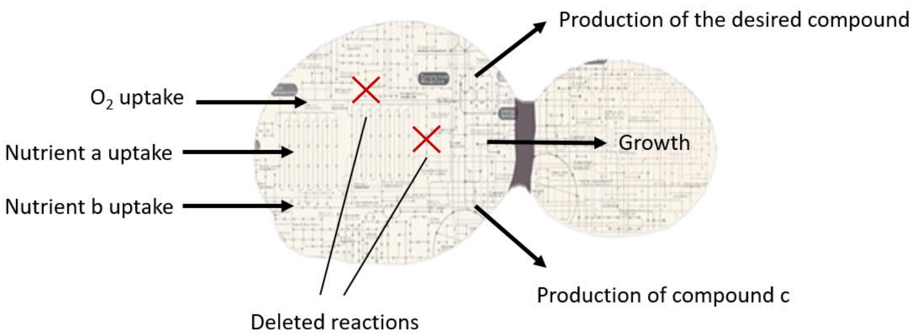

Figure 4. When a reaction is deleted the metabolic fluxes have to redistribute. Therefore, strains can be improved to achieve a desired phenotype e.g. by deleting reactions in a way that more flux is directed to product formation. Reaction deletions can be achieved by gene deletions according to gene-protein-reaction rules. Cell figure template obtained from Paula Jouhten.

2.2.3 Algorithms for designing host improvement strategies

Several algorithms that use GEMs for designing engineering strategies for improving production levels have been developed. Some of these algorithms aim at coupling the production with growth (growth-product coupling), and others try to find push- and pull-strategies to increase production. Push- and pull-strategies include engineering targets for increasing the activity of the product synthesis pathway (i.e. pull) and/or for increasing the flux towards the product precursors (i.e. push).

To create growth-product coupling the engineering strategy includes such modifications to the metabolic network that the formation of the desired product becomes a byproduct of growth. Different approaches have been taken for identifying such strategies. OptKnock algorithm solves a bilevel optimization problem (Bugard *et al.*, 2003). The outer problem finds metabolic reaction deletions (achievable with gene deletions) that maximize the production of the compound of interest while the inner problem maximizes the cellular objective (i.e. growth) subject to the reaction deletions (Figure 5). OptKnock has been implemented as a mixed-integer linear programming (MILP) problem that has both continuous and binary variables.

While OptKnock identifies strategies yielding highest possible production, the successor algorithms RobustKnock (Tepper and Shlomi, 2010) and EMILiO (Yang e*t al.*, 2011) search for engineering designs leading to highest necessary production per growth (i.e. maximize the minimal necessary production rate/yield). Another difference between OptKnock and the extension algorithms to OptKnock is that EMILiO and OptReg (Pharkya and Maranas, 2006) suggest in addition to gene deletion targets also up- and downregulation targets. However, it is good to remember that strong growth-product coupling in real life may be created only with gene deletions as the cells might lose the changes in the flux regulation faster than replacing gene deletions due to natural selection. OptStrain (Pharkya *et al.*, 2004) and SimOptStrain (Kim *et al.*, 2011) combine growth-coupling with searching of genes from different species to find the least amount of non-native genes that has to be introduced in the production host to achieve as good as possible heterologous production of a desired compound.

maximize bioengineering objective
(through gene knockouts)

subject to

> **maximize cellular objective**
> (over fluxes)
>
> **subject to**
> - fixed substrate uptake
> - network stoichiometry
> - biomass production ≥ predefined minimal biomass production
> - ATP production ≥ non-growth-associated ATP maintenance requirement
> - blocked reactions identified by outer problem

number of knockouts ≤ limit

Figure 5. Bilevel optimization framework of OptKnock. Figure modified from Bugard *et al*. (2003).

OptGene (Patil *et al.*, 2005) is an example of a metaheuristic growth-coupling optimization algorithm. Metaheuristic algorithms make use of randomness, but have also a strategy guiding the optimization. While deterministic algorithms always produce the same output, metaheuristic algoritms might instead of global optimal solution converge to a local optimal solution (Maia *et al*., 2016). However, compared to deterministic algorithms, metaheuristic algorithms are usually faster and allow solving problems of larger size. In addition, in contrast to linear programming metaheuristic algorithms allow using also nonlinear objective functions (Patil *et al.*, 2005). OptGene uses a genetic algorithm for the optimization, to find the best gene knock-out strategy to obtain a desired phenotype (Patil *et al.*, 2005). In OptGene FBA, MOMA or ROOM phenotype prediction algorithms can be used as the fitness function in the genetic algorithm (Patil *et al.*, 2005).

Both OptKnock and OptGene have been used to find metabolic network modifications for improving production of different compounds. OptKnock has been used to find gene deletion strategies for growth-product coupling *in silico* for several different compounds including amino acids (Pharkya *et al*., 2003), lactate (Bugard *et* al., 2003; Feist *et al*., 2010), succinate (Bugard *et* al., 2003; Feist *et al*., 2010), ethanol (Feist *et al*., 2010), glycerol (Feist *et al*., 2010), 1,3-propanediol (Bugard e*t al*., 2003), pyruvate (Feist *et al*., 2010), fumarate (Feist *et al*., 2010), 2-oxoglutarate (Feist *et al*., 2010) and lipids (Castañeda *et al*., 2019). In addition, knockouts predicted by OptKnock have been shown to increase e.g. lactate (Fong *et al*., 2005) and succinate (Suthers e*t al*., 2020) production *in vivo*. Asadollahi *et al*., (2009) used OptGene to design gene deletions for improved sesquiterpene production, Brochado *et al*., (2010) used OptGene to improve vanillin production, and Iranmanesh *et al.* (2020) defined gene deletions with OptGene to improve L-phenylacetylcarbinol production. In all of these cases improved production was validated *in vivo*.

However, even if growth-product coupling strategies have been designed for many compounds not all products can be growth-coupled (Campodonico *et al.*, 2014; Feist *et al.*, 2010). OptForce (Ranganathan *et al.,* 2010), CosMos (Cotten and Reed*,* 2013) and FVSEOF (Park e*t al*., 2012) are algorithms that instead of finding ways to couple the production with growth aim at increasing the product yield with gene deletions, downregulations and upregulations in a way that higher flux is directed to product formation (i.e. push strategy). Xu *et al.* (2011) were able to increase intracellular malonyl-CoA levels 4-fold and then direct this to heterologous naringenin production by performing gene deletions and gene upregulations suggested by OptForce. Ranganathan *et al*. (2012) used OptForce to increase production of fatty acids. In addition, OptForce has been used *in silico* to find suitable network modification strategies for overproduction of succinate (Ranganathan *et al.,* 2010) and for heterologous production of 6-deoxyerythronolide B, erythromycin D, yersiniabactin, and salicylate 2-O-β-D-glucoside (Moscatello and Pfeifer, 2018). However, flux regulation is complex, and therefore some of the solutions suggested by algorithms that include adjusting flux levels might be difficult to realize *in vivo.* For instance, engineering attempts focusing on gene expression levels may not lead into flux change as the flux can be regulated post-transcriptionally or post-translationally, or the flux control of the pathway is on other enzymes. This applies also to the growth-product coupling strategies that introduce flux up- or downregulation.

2.3 Adaptive laboratory evolution (ALE) for host improvement

A way to circumvent the need to know how to overcome the native flux regulation is to use ALE for strain improvement. ALE is easy to perform and no knowledge about an organisms genotype-phenotype relations is needed (Dragosits and Mattanovich, 2013). In ALE, the cells are cultured for a prolonged time in a chosen environment. During ALE random mutations arise (either naturally or by mutagenesis) and mutations substantially beneficial for fitness in the particular chosen condition are enriched due to Darwinian selection (Dragosits and Mattanovich, 2013; Sandberg *et al.*, 2019). However, it is possible to use ALE only when the desired trait is aligned with cellular fitness. A desired production pathway is rarely natively coupled with growth. Strains overproducing the product could be obtained by screening of variants that have emerged, but such artificial selection is limited by the number of variants that can be generated and the throughput of measuring their production performance.

The specific advantage with growth-product coupling is that the desired production can be improved together with fitness using ALE (Fong *et al.*, 2005). As discussed above, such growth-product coupling can be achieved with suitable metabolic network modifications. Another possibility to couple target reactions with growth is by choosing suitable environmental conditions. Reyes *et al*. (2014) coupled the production of carotenoids having antioxidant properties with fitness by using hydrogen peroxide shocks during ALE. Nutritional environment also affects how the flux is distributed in the metabolic network for achieving optimal cellular fitness (Pál and Papp, 2017; Szappanos *et al*., 2016). Therefore, the chemical environment of cells can be designed to determine which metabolic fluxes are growth-coupled (i.e. have non-zero response to natural selection) (Jouhten and Patil, 2016; Jouhten *et al.*, 2021 Preprint). Intuitive design of a suitable nutritional environment for specific reactions is nonetheless very difficult if not impossible.

To tackle this, an algorithm (EvolveX) that uses GEMs for identifying nutritional environments (selection niche) that create dependencies between desired target fluxes (e.g. reactions of a production pathway) and fitness (i.e. growth), was recently developed (Jouhten & Patil, 2016; Jouhten *et al.*, 2021 Preprint). Therefore, when the designed environment is used, the mutations that enrich in ALE are expected to increase the flux in these target reactions (or decrease as the desired direction of change can be selected). The mutations in the genomes of evolved clones are then expected to enhance the particular reaction fluxes even when the clones are transferred to another environment, the target niche. When aiming to increase production of a desired product, reactions of the production pathway can be chosen as up regulation targets.

Using a selection niche (i.e. specific environmental conditions) for coupling specific metabolic fluxes to growth in ALE allows desired, not necessarily fitness increasing, trait improvement also when the use of genetic engineering is not possible e.g. in food applications (Jouhten e*t al.*, 2021 Peprint). In industrial biotechnological applications also genetic engineering is possible. Therefore, Paula Jouhten designed an algorithmic extension of EvolveX (EvolveXGA) that search suitable combinations of selection niche and metabolic network modifications to couple selected target reactions with growth.

2.4 EvolveXGA - model guided design of ALE strategies for host improvement

EvolveXGA is a metaheuristic algorithm where EvolveX has been implemented as a fitness function into a genetic algorithm. The genetic algorithm (Figure 6) can vary both the selection niche and metabolic reaction deletions and the EvolveX algorithm as the fitness function scores the combinations. The aim of EvolveXGA is to define a suitable selection niche and pre-ALE reaction deletions in a way that the fluxes of selected target reactions are under as strong as possible selection during ALE. Therefore, the strain can be improved through Darwinian selection using ALE and mutations that increase the flux in the target reactions are likely to enrich.

After ALE the evolved clones can be turned into production strains or the mutations that have enriched during ALE and improved the target fluxes, can be identified with genome sequencing and introduced to the background strain with the production pathway. It would also be possible to determine which enriched mutations were causal for fitness increase and improve production, and which mutations are hitchhikers. Hitchhiker mutations are neutral or deleterious mutations that are co-enriched with beneficial mutations (Smith and Haigh, 1974). They can become enriched due to co-localization (i.e. by being closely located in the genome to a gene with a beneficial mutation) or just by chance if having been in the clone that gained the beneficial mutation.

Assumption is that the beneficial mutations that have been enriched during ALE are not lost during production. However, as production with the evolved clones is performed in the target environment which differs from the selection niche used during ALE the switch of niche may release the growth-coupling, and the flux in the target reactions is not selected anymore. Therefore, it is possible that the improved flux in the selected target reactions could be lost. However, as the actual applications (e.g. production of the desired compound) are usually short in the number of generations compared with the generation number used in ALE (Jouhten *et al.*, 2021 Preprint), the chance for variants that cause decrease in the target fluxes taking over the process is low. In addition, since the selection niche contains simple nutrients, the particular selection niche environment could be used for biomass production and maintenance before producing the desired compound in the target niche.

To turn the evolved clones into production strains the improved flux of the target reactions has to be directed to product formation. This can be done by allowing the growth-essential biomass component synthesis, that has been possible only through the target reactions during ALE, through an alternative route (e.g. by weakly expressing some of the deleted genes). This would cause push of resources towards production as the flux of the target reactions is not anymore used for growth.

2.5 The aim of the study

The aim of this work was to test and evaluate EvolveXGA algorithm for finding strategies to improve small molecule production by a microbial host using adaptive laboratory evolution. The EvolveXGA algorithm had been developed by Paula Jouhten and the initial python implementation had been made by Sandra Castillo. In this work the EvolveXGA algorithm was tested and evaluated using yeast *S. cerevisiae* as a model microbial host.

First the algorithm was set up by defining variable values of the genetic algorithm (e.g. generation number, size of population, amount of nutrients in the nutritional environment of adaptive evolution and the number of reaction deletions). Next EvoleXGA was used to search reaction deletion strategies to couple the reactions of oxalate pathway (a heterologous pathway for glycolic acid synthesis) with growth and the results were compared with an intuitive solution designed by Paula Jouhten. Then, combined reaction deletion and nutritional environment strategies were searched to growth-couple the oxalate pathway reactions, and finally the generalizability of EvolveXGA was tested by searching growth-coupling strategies for 28 other heterologous production pathways.

3 Materials and methods

3.1 Genome-scale metabolic model

The *S. cerevisiae* consensus genome-scale model (Lu *et al.*, 2019), v. 8.3.5 (Sánchez *et al.,* 2020) was used with some modifications to flux lower and upper bounds (APPENDIX 1). Oxalate pathway for glycolic acid synthesis (Table 1) and other heterologous pathways (APPENDIX 2) were individually introduced to the model. All these changes were made using Jupyter notebook v. 6.0.3, python v. 3.7.6 and cobra py v. 0.18.1 https://github.com/opencobra/cobrapy.

3.2 EvolveX algorithm

EvolveX algorithm (Jouhten and Patil, 2016; Jouhten *et al.*, 2021 Preprint) was used to score the individuals in the genetic algorithm for the growth-coupling of defined target flux(es). While EvolveX is compatible with input including both up- and downregulation target reactions, here only upregulation target reactions were considered (i.e. the aim was to increase the flux in the selected target reactions). In EvolveX, first the minimal total nutrient uptake required for a certain unit of biomass formation is determined given the particular selection niche and metabolic network (Equation 4).

$$\min \boldsymbol{a} \cdot \boldsymbol{v} \qquad \text{(Equation 4)}$$

s.t.

$$\boldsymbol{S} \cdot \boldsymbol{v} = 0$$

$$v_{\mu} = b$$

$$\boldsymbol{v_{del}} = 0, \qquad del \in D$$

$$\boldsymbol{v_{lb}} \leq \boldsymbol{v} \leq \boldsymbol{v_{ub}}$$

where $\boldsymbol{a}$ defines the possible nutrient uptakes (i.e. glucose and ammonium, when only gene-deletion strategies where searched and selection niche nutrients when combined suitable gene deletions and selection niche was searched.), v_{μ} is flux of biomass production, b was set to an arbitrary value of 10 when only optimal gene deletion strategies were searched and to 1 when gene deletion strategies were combined with optimal selection niche, D is the set of metabolic reactions that form the gene deletion strategy, $\boldsymbol{v_{lb}}$ and $\boldsymbol{v_{ub}}$ are flux lower and upper bounds respectively.

Next the minimum sum of the target fluxes relative to growth (i.e. fitness) was determined (Equation 5) and this formed the fitness score for each individual. This score represents the total growth-coupling of the target fluxes.

$$\min \sum_{|u|=1}^{k} v_{|u|} \qquad \text{(Equation 5)}$$

s.t.

$$\boldsymbol{S} \cdot \boldsymbol{v} = 0$$

$$v_{\mu} = b$$

$$v_{del} = 0$$

$$\boldsymbol{v_{lb}} \leq \boldsymbol{v} \leq \boldsymbol{v_{ub}}$$

$$\boldsymbol{a} \cdot \boldsymbol{v} \geq r_{uptake,max}$$

$$0 \leq v_{|u|} \leq v_{|u|,ub}, \qquad |u| \in U$$

$$v_u - v_{|u|} \leq 0, \qquad u \in T$$

$$-v_u - v_{|u|} \leq 0$$

where $r_{uptake,max}$ is the optimal nutrition uptake (Equation 4), T is the set of upregulation targets, U is the set of absolute flux variables of these targets and $v_{|u|,ub}$ are the upper bounds of the absolute flux variables.

3.3 Genetic algorithm

Sandra Castillo had performed an initial implementation of EvolveXGA algorithm by introducing EvolveX algorithm implemented in python as a fitness function in pyeasyga genetic algorithm (https://github.com/remiomosowon/pyeasyga). This python implementation was used as a starting point for the development of the genetic algorithm in EvolveXGA. EvolveXGA was run with python v. 3.7.6 for simulations of the model containing the glycolic acid production pathway and with python v. 3.6.8 when simulations were performed for other heterologous pathways. IBM ILOG CPLEX v. 12.10 was used as a linear programming (LP) and mixed-integer linear programming (MILP) problem solver.

3.4 SwitchX algorithm

The minimum set of fluxes requiring up- or downregulation (compared with the wild type phenotype fluxes) to improve the production of a desired compound were determined by solving a MILP problem. The MILP problem was formulated in a similar way as ROOM (Shlomi e*t al*., 2005) (Equation 6). Then fluxes requiring upregulation were extracted.

$$\min \sum_{i=1}^{m} y_i \qquad \text{(Equation 6)}$$

s.t.

$$\boldsymbol{S} \cdot \boldsymbol{v} = 0$$

$$\boldsymbol{v_{lb}} \leq \boldsymbol{v} \leq \boldsymbol{v_{ub}}$$

$$v_{production} \geq \alpha \cdot v_{production,max} \qquad \alpha \in [0\ 1]$$

$$v_i - y_i\left(v_{ub,i} - w_{ub,i}\right) \leq w_{ub,i}$$

$$v_i - y_i\left(v_{lb,i} - w_{lb,i}\right) \geq w_{lb,i} \qquad y_i \in [0,1]$$

$$w_{ub,i} = w_{max,i} + \delta\left|w_{max,i}\right| + \varepsilon$$

$$w_{lb,i} = w_{min,i} - \delta\left|w_{min,i}\right| - \varepsilon$$

where y_i are binary variables defining whether a reaction flux i has to be up- or downregulated for the production to reach a level α of maximum production. $w_{ub,i}$ and $w_{lb,i}$ are threshold fluxes of the wildtype phenotype. $w_{min,i}$ and $w_{max,i}$ are wildtype minimum and maximum fluxes respectively. α, δ and ε are parameters. In this work α was given a value 0.99, δ had a value 0.5, and ε had a value 0.00001.

4 Results and discussion

4.1 EvolveXGA algorithm

EvolveXGA is an optimization algorithm for designing strategies to couple selected target reactions with growth. Therefore, EvolveXGA can be used to define strategies that allow the use of ALE for improving production by microbial hosts. EvolveXGA allows searching strategies for growth-coupling of metabolic reactions using both metabolic network modifications and suitable selection niche. The earlier methods have enabled either identifying metabolic network modifications (Patil *et al.*, 2005) or identifying a suitable growth environment (Jouhten and Patil, 2016; Jouhten *et al.*, 2021 Preprint). As OptGene (Patil *et al.*, 2005), EvolveXGA algorithm is implemented as a genetic algorithm (Figure 6). The genetic algorithm performs optimization by mimicking Darwinian evolution, and therefore, the nomenclature of genetic algorithms also originates from evolution. First, an initial "population" is formed by creating "individuals" consisting of a "chromosome" that represents a candidate solution. In EvolveXGA the candidate solution is the reaction deletion and selection niche strategy to achieve growth-coupling. Then the "fitness" of the individuals is calculated in a suitable way for the particular optimization task. While OptGene was demonstrated for using e.g. FBA or MOMA as fitness functions, EvolveXGA integrates recently developed EvolveX (Jouhten and Patil, 2016; Jouhten *et al.*, 2021 Preprint) as a fitness function. The fitness of the individuals is then improved by performing "mutations" and "crossovers" and a new population is formed from the improved individuals as well as from the best individuals of the previous generation. The individuals of the new population are then scored and this loop of forming new individuals and scoring them is continued until a predefined number of generations is reached. The individual with the highest score represents the reaction deletion and selection niche strategy that creates the best identified coupling of the selected target reactions with growth. A genetic algorithm does not necessarily find the globally optimal solution even with high numbers of generations. The details of EvolveXGA implementation is discussed below.

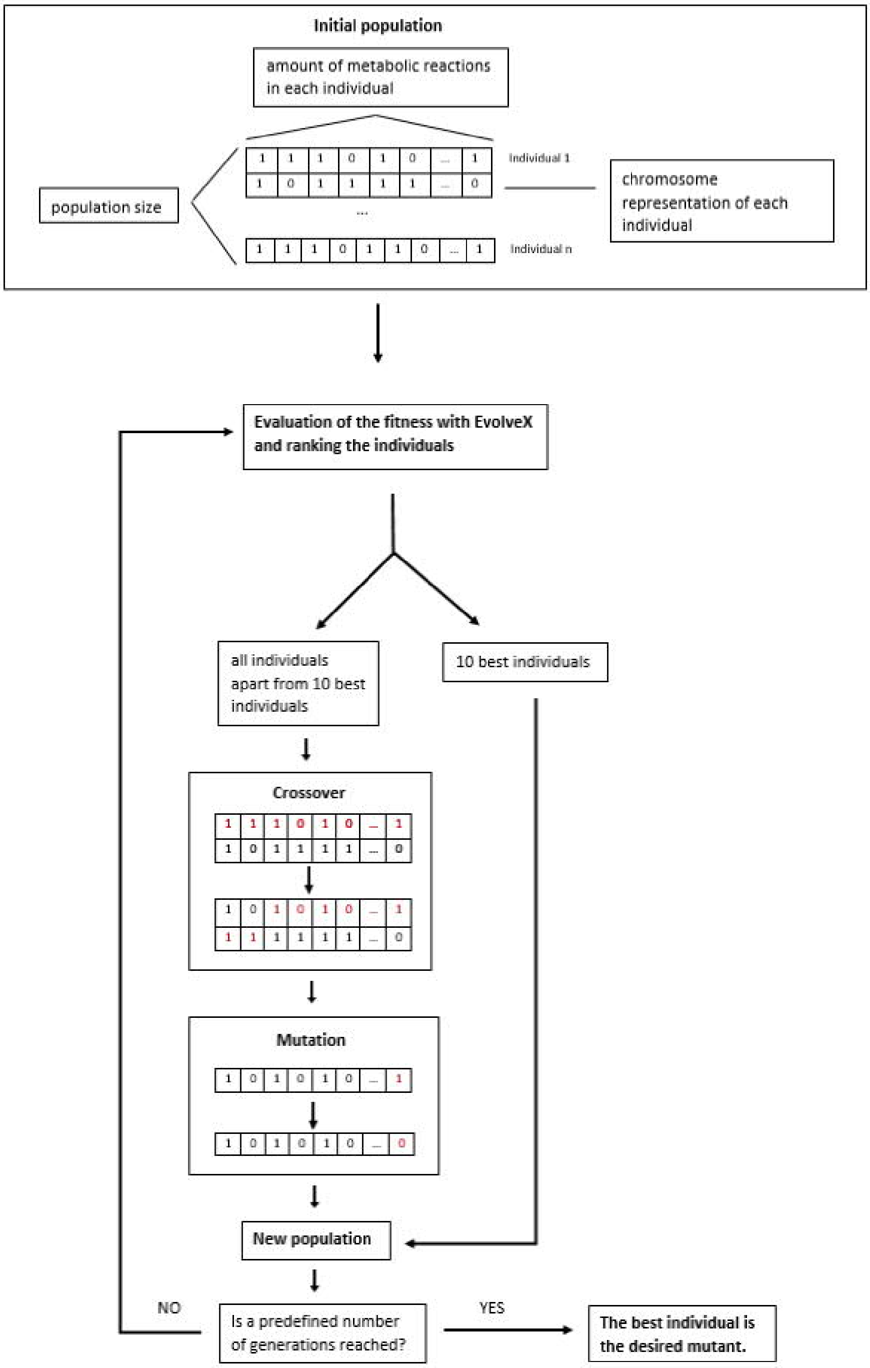

Figure 6. Schematic figure of EvolveXGA algorithm. In genetic algorithms a set of individuals form the population. Each individual is presented with a chromosome which consist of all metabolic reactions either present (1) or deleted (0). After forming initial population each individual is scored with EvolveX. The ten best individuals are selected directly to the new population. Rest of the individuals are selected for crossover and mutation and the new population is filled with the new individuals until the original population size is reached. Then

the new population is again scored with EvolveX. This is continued until a predefined amount of generations is reached. Figure modified from Patil *et al*. (2005).

4.1.1 Search spaces of reaction deletions and selection niche components

EvolveXGA defines reaction deletions to achieve growth-coupling from the reactions in the GEM. However, if the search space would consist of all the reactions in the GEM it would reduce the probablity of finding the best solution or at least increase the calculation time remarkably. In addition, deleting reactions essential for growth would be detrimental for the cell. Therefore, the search space was limited by removing several reactions from possible reaction deletion targets. Essential reactions both for growth and for the heterologous product synthesis in question were determined with FVA and removed. Also reactions that carried no flux under any nutritional environment (i.e. blocked reactions) were determined with FVA and removed. In addition, reactions without gene annotation; transport, exchange and diffusion reactions; and some additional reactions e.g. reactions taking part to lipid metabolism (APPENDIX 3) were removed. Furthermore, the selection niche was searched from a limited list of 37 carbon and nitrogen sources (APPENDIX 4).

4.1.2 "Chromosome" formed of metabolic reactions and selection niche components

A binary variable was assigned to each reaction that was considered a candidate for removal by gene deletion(s) and to each nutrient that could be a component of the selection niche. Binary variable value 0 for a reaction indicated that the corresponding reaction flux was restricted to 0 (i.e. reaction deletion). For convenience, the nutrients with binary variable value 0 formed the selection niche. Thus, the set of these variables (i.e. chromosome in the terminology of genetic algorithms) formed an individual in the EvolveXGA algorithm (Figure 6).

4.1.3 Initialization of the "population"

A population of individuals (Figure 6) is created to initialize the optimization using a genetic algorithm. Patil *et al*. (2005) tested different population sizes for OptGene and according to them better results were achieved when the population size increased although no significant increase in the results were achieved when population size was beyond 125 individuals. However, according to Sarker and Kazi (2003) it is necessary to increase the population size when the search space increases. As Patil *et al*. (2005) did not specify their search space, a slightly bigger population size of 200 individuals was used in this work to ensure that finding optimal solutions was not restricted by it. The initial population was created by forming individuals until the population size was reached. Each individual was set to contain four random reaction deletions and a selection niche with three randomly selected nutrients (i.e. carbon and nitrogen sources). The number of reaction deletions and selection niche components in an individual was kept constant in the optimization.

Patil *et al*. (2005) defined gene deletion strategies with OptGene including three to six gene deletions and the yields increased when the number of gene deletions increased. However, increasing the number of gene deletions increases also the probability that it would not be possible to realize some of the gene deletions *in vivo* due to effects on cell viability. In addition, increasing the number of gene deletions increases the solution space exponentially, and therefore, the genetic algorithm will search a smaller fraction of the total solution space during one search (Patil e*t al*., 2005). Higher number of nutrients in the selection niche increases possibility that complex growth dynamics beyond the predictive ability of GEMs will take place due to nutrient preferences of microbes.

4.1.4 Scoring the "individuals"

After initializing the population every individual was scored with EvolveX algorithm (Methods, Equation 4 and 5) (Jouhten and Patil, 2016; Jouhten *et al*., 2021 Preprint). The score attained with EvolveX is the minimum sum of the target reaction fluxes relative to growth, and therefore, represents the total growth-coupling of the target reactions. EvolveX used the *S. cerevisiae* GEM from which reactions that were deleted in the individual were removed. In addition, EvolveX used those nutrients that were set to the selection niche for each individual.

4.1.5 Forming new "population"

After scoring all the individuals a new population was formed by elitism, crossover and mutation. Ten best individuals of the previous generation were directly selected to the next population (i.e. elitism). Then individuals for crossover and mutation were selected randomly from the previous generation and the new generation was filled with the new individuals. Another possibility could have been to select individuals for crossover and mutation with Roulette wheel method that is used in OptGene (Patil *et al*., 2005). In Roulette wheel method the fitness score of an individual affects how likely it will be chosen. When the score is higher the probability to become selected is higher.

Crossovers and mutations were performed with a probability of 0.8 for both. In crossover two randomly selected individuals from the previous generation ("parents") were crossed (i.e. new combinations of the parents were formed) to produce new individuals ("children"). Crossover was performed separately for reactions and nutrients. First all the reactions with value 0 from the two parents were collected on a list. Then all possible four reaction combinations from these reactions were determined. Finally, new children were created by selecting randomly one of these combinations for each children. When using EvolveXGA for finding combined reaction deletion and selection niche strategies the same crossover procedure was performed also for the nutrients with the change that now three nutrient combinations were determined and chosen.

For creating a mutation these children from crossover were mutated. First one of the reactions that had value 0 was randomly selected and set to 1. After this another reaction was randomly selected from all the reactions that had value 1 and set to 0.

The same was performed with compounds when in addition to reaction deletions also a suitable selection niche was defined.

4.1.6 Termination of optimization

The optimization of the reaction deletion and selection niche strategies was let to continue until a predefined amount of 1000-5000 generations was reached. The amount of variables affect how fast the fitness score reach the optimal solution. Therefore, as expected the fitness score continued to increase for a higher number of generations when combined reaction deletion and selection niche strategies were searched than when only reaction deletions were defined (Figure 7). However, in both cases the EvolveXGA reached a plateau well before 1000 generations.

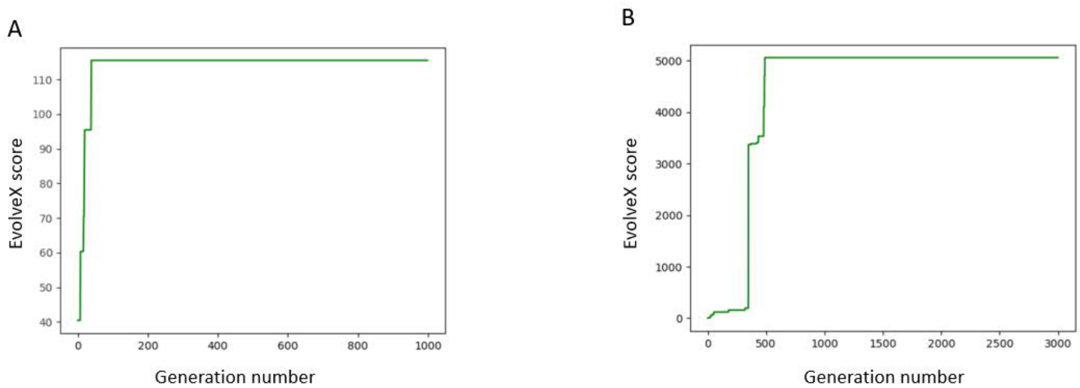

Figure 7. Convergence curves of EvolveXGA. A) Convergence curve example when searching only reaction deletions to couple glycolic acid synthesis pathway reactions with growth . B) Convergence curve example when searching in addition to reaction deletions also a suitable selection niche to couple glycolic acid synthesis pathway reactions with growth. The convergence curves were similar to the convergence curve that Patil *et al*., (2005) presented for OptGene algorithm.

4.2 Reaction deletion strategies for improving glycolic acid production using adaptive laboratory evolution

Glycolic acid can be produced using *S. cerevisiae* with a heterologous oxalate pathway consisting of oxaloacetase, oxalate-CoA ligase, glyoxylate dehydrogenase and glyoxylate reductase (Toivari *et al.*, 2017) (Table 1). The oxalate pathway for glycolic acid production has high yield and low oxygen requirement. First, it was tested if EvolveXGA could find such reaction deletions that the three first reactions of the oxalate pathway (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) would become growth-coupled and selected during ALE. In this first stage glucose and ammonium were used as the only carbon and nitrogen sources, respectively (i.e. minimal growth medium).

Table 1. Reactions of the oxalate pathway for glycolic acid synthesis.

reaction name	reaction
Oxaloacetase	oxaloacetate [cytoplasm] + H2O [cytoplasm] --> acetate [cytoplasm] + oxalate [cytoplasm]
Oxalate-CoA ligase	oxalate [cytoplasm] + CoA [cytplasm] + ATP [cytoplasm --> oxalyl-CoA [cytoplasm] + AMP [cytoplasm] + diphosphate [cytoplasm]
Glyoxylate dehydrogenase	oxalyl-CoA [cytoplasm] + NADPH [cytoplasm] + H+ [cytoplasm] --> CoA [cytoplasm] + glyoxylate [cytoplasm] NADP(+) [cytoplasm]
Glyoxylate reductase	glyoxylate [cytoplasm] + NADPH [cytoplasm] + H+ [cytoplasm] --> glycolate [cytoplasm] + NADP(+) [cytoplasm]
Glycolate transport	glycolate [cytoplasm] --> glycolate [extracellular]
Glycolate exhange	glycolate [extracellular] -->

An intuitive solution (Figure 8) to achieve growt-coupling of the first three reactions of oxalate pathway for glycolic acid synthesis (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) had already been derived by Paula Jouhten from the solution presented by Otero *et al*. (2013) to couple succinate production with growth. In the intuitive solution, native pathways to produce glycine and serine are blocked, and therefore, the cell has to produce these by using near-equilibrium threonine aldolase or a heterologous oxalate pathway. The reaction deletion strategies that were identified using EvolveXGA where compared against this intuitive design.

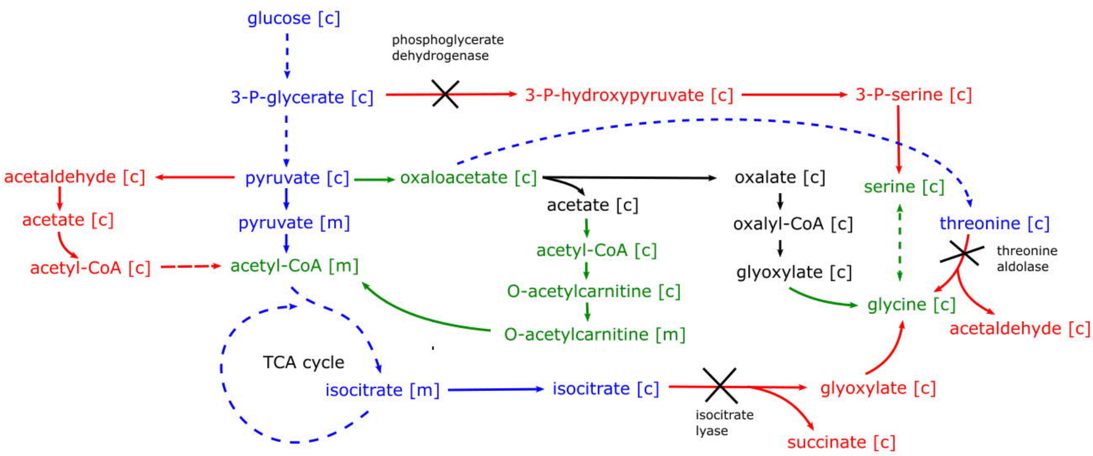

Figure 8. Intuitive solution to couple first three reactions of oxalate pathway for glycolic acid synthesis (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) with growth. Deletion of phosphoglycerate dehydrogenase forces serine to be produced from glycine. Deletion of isocitrate lyase and threonine aldolase forces glycine to be produced through the oxalate pathway. Even if pyruvate decarboxylase is not deleted model simulations do not predict flux through it as it is more beneficial for growth yield on available nutrients to produce acetate through the oxalate pathway. Serine can be formed from glycine either with cytosolic glycine hydroxymethyltransferase or mitochondrial glycine hydroxymethyltransferase. Blue, red and green colours refer to native *S. cerevisiae* metabolism and black reactions are heterologous oxalate pathway reactions. Red colour refers to interrupted reactions and green colour to upregulated reactions. Solid line refers to one reaction and dashed line to multiple reactions.

4.2.1 EvolveXGA solutions of reaction deletions

As genetic algorithms do not search the whole search space it is possible that the algorithm finds different locally optimal solutions in independent runs. Therefore, the algorithm was run ten times for 1000 generations to see if different solutions would be obtained on different runs. However, only two final solutions were found. Solution A was attained nine times out of ten and solution B once (Table 2). Nevertheless, in addition to these final solutions the best solution of every generation was collected and these solutions included several different growth-coupling strategies for the reactions of the oxalate pathway for glycolic acid synthesis.

Table 2. Three different solutions to couple the three first reactions of oxalate pathway for glycolic acid synthesis (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) with growth. Solutions A and B are attained with EvolveXGA and the third solution is an intuitive solution designed by Paula Jouhten.

Solution A				
EvolveX score	11.55 mmol/g CDW			
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0113	acetyl-CoA synthetase	acetate [mitochondrion] + ATP [mitochondrion] + coenzyme A [mitochondrion] --> acetyl-CoA [mitochondrion] + AMP [mitochondrion] + diphosphate [mitochondrion]	*ACS1*
	r_0658	isocitrate dehydrogenase (NAD+)	isocitrate [mitochondrion] + NAD [mitochondrion] --> 2-oxoglutarate [mitochondrion] + carbon dioxide [mitochondrion] + NADH [mitochondrion]	*IDH1*, *IDH2*
	r_0959	pyruvate decarboxylase	H+ [cytoplasm] + pyruvate [cytoplasm] --> acetaldehyde [cytoplasm] + carbon dioxide [cytoplasm]	*PDC1*, *PDC5*, *PDC6*
	r_0961	pyruvate dehydrogenase	coenzyme A [mitochondrion] + NAD [mitochondrion] + pyruvate [mitochondrion] --> acetyl-CoA [mitochondrion] + carbon dioxide [mitochondrion] + NADH [mitochondrion]	*PDB1*, *PDA1*, *LPD1*, *PDX1*, *LAT1*
Solution B				
EvolveX score	5.08 mmol/g CDW			
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0032	3',5'-bisphosphate nucleotidase	adenosine 3',5'-bismonophosphate [cytoplasm] + H2O [cytoplasm] --> AMP [cytoplasm] + phosphate [cytoplasm]	*MET22*
	r_0662	isocitrate lyase	isocitrate [cytoplasm] --> glyoxylate [cytoplasm] + succinate [cytoplasm]	*ICL1*
	r_0891	phosphoglycerate dehydrogenase	3-phosphonato-D-glycerate(3-) [cytoplasm] + NAD [cytoplasm] --> 3-phospho-hydroxypyruvate [cytoplasm] + H+ [cytoplasm] + NADH [cytoplasm]	*SER3*, *SER33*
	r_1040	threonine aldolase	L-threonine [cytoplasm] --> acetaldehyde [cytoplasm] + L-glycine [cytoplasm]	*GLY1*
Intuitive solution				
EvolveX score	4.49 mmol/g CDW			
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0662	isocitrate lyase	isocitrate [cytoplasm] --> glyoxylate [cytoplasm] + succinate [cytoplasm]	*ICL1*
	r_0891	phosphoglycerate dehydrogenase	3-phosphonato-D-glycerate(3-) [cytoplasm] + NAD [cytoplasm] --> 3-phospho-hydroxypyruvate [cytoplasm] + H+ [cytoplasm] + NADH [cytoplasm]	*SER3*, *SER33*
	r_1040	threonine aldolase	L-threonine [cytoplasm] --> acetaldehyde [cytoplasm] + L-glycine [cytoplasm]	*GLY1*

The best solution suggests that the genes encoding mitochondrial acetyl-CoA synthetase, mitochondrial NAD^+ dependent isocitrate dehydrogenase, pyruvate decarboxylase and pyruvate dehydrogenase should be deleted to couple the first three reactions of the oxalate pathway for glycolic acid synthesis with growth as strongly as possible (i.e. as high as possible flux required per growth flux). However, the deletion of mitochondrial acetyl-CoA synthetase contributes very little to the final EvolveX score (the score with acetyl-CoA synthetase deleted is only 0.000024 mmol/g CDW bigger than the score without deleting acetyl-CoA synthetase when the score with all the deletions is 11.55 mmol/g CDW).

Deleting pyruvate decarboxylase, pyruvate dehydrogenase and mitochondrial acetyl-CoA synthetase removes other possibilities to produce acetyl-CoA, and therefore, acetyl-CoA has to be produced by cytoplasmic acetyl-CoA synthetase from acetate generated by oxaloacetase. This acetyl-CoA can then be transported to mitochondria by carnitine/acetyl-carnitine shuttle (Krivoruchko *et al.*, 2015) (Figure 9). However, deleting mitochondrial acetyl-CoA synthetase does not increase the score notably as its predicted flux is very low. When FVA was performed with pyruvate decarboxylase, pyruvate dehydrogenase and NAD^+ dependent isocitrate dehydrogenase deleted the flux through mitochondrial acetyl-CoA synthetase relative to growth was only 8.0 x 10^{-6} mmol/g CDW. Deleting pyruvate decarboxylase and pyruvate dehydrogenase decreases also the use of pyruvate, and therefore, more pyruvate can be converted to oxaloacetate by pyruvate carboxylase (Figure 9).

When mitochondrial NAD^+-dependent isocitrate dehydrogenase is deleted the cell has to use the $NADP^+$-dependent isocitrate dehydrogenase and therefore NADPH is formed. This is beneficial for the oxalate pathway activity as glyoxylate dehydrogenase requires NADPH (Table 1). Therefore, with the deletions of solution A oxalate pathway becomes the only possible pathway to produce acetyl-CoA and, in addition, not the only possible but the most beneficial pathway to produce glycine and serine for growth.

Solution B (Figure 10) includes the same deletions as the intuitive solution (Figure 8), but, in addition, 3',5'-bisphosphate nucleotidase is deleted. Adenosine 3',5'-bismonophosphate is an obligatory byproduct of homocysteine synthesis (Thomas and Surdin-Kerjan, 1997), and therefore, its formation is necessary for growth. Adenosine 3',5'-bismonophosphate is further transformed to AMP and phosphate. AMP is needed in RNA biosynthesis. Also acetyl-CoA synthetase and oxalate-CoA ligase produce AMP. When 3',5'-bisphosphate nucleotidase is deleted adenosine 3',5'-bismonophosphate cannot be converted to AMP and the acetyl-CoA synthetase and oxalate-CoA ligase fluxes are increased. The deletion appears to have interrupted the most beneficial way to produce AMP leaving the AMP that is produced through acetyl-CoA synthetase and oxalate-CoA ligase to be used for RNA biosynthesis (Figure 10).

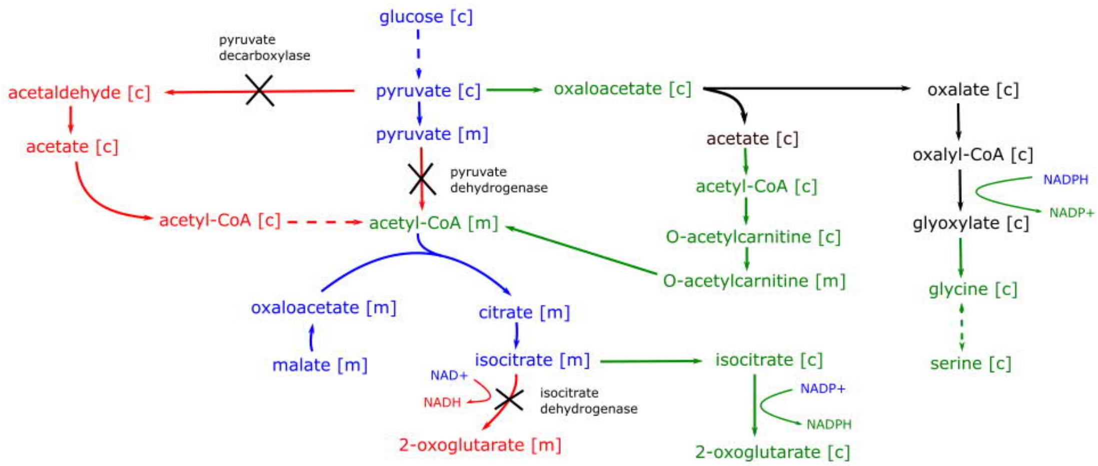

Figure 9. The solution with highest EvolveX score to couple three first oxalate pathway reactions for glycolic acid synthesis (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) with growth. Deletion of pyruvate decarboxylase and pyruvate dehydrogenase forces acetyl-CoA to be produced through oxaloacetase, the first reaction of oxalate pathway. Deletion of NAD^+ dependent isocitrate dehydrogenase increases the flux in the $NADP^+$ dependent isocitrate dehydrogenase increasing NADPH generation. In addition, the deletions make oxalate pathway the most favorable way to produce glycine and serine. Serine can be formed from glycine either with cytosolic glycine hydroxymethyltransferase or mitochondrial glycine hydroxymethyltransferase. Mitochondrial acetyl-CoA synthetase is omitted from the picture. Blue, red and green colours refer to native *S. cerevisiae* metabolism and black reactions are heterologous oxalate pathway reactions. Red colour refers to interrupted reactions and green colour to upregulated reactions. Solid line refers to one reaction and dashed line to multiple reactions.

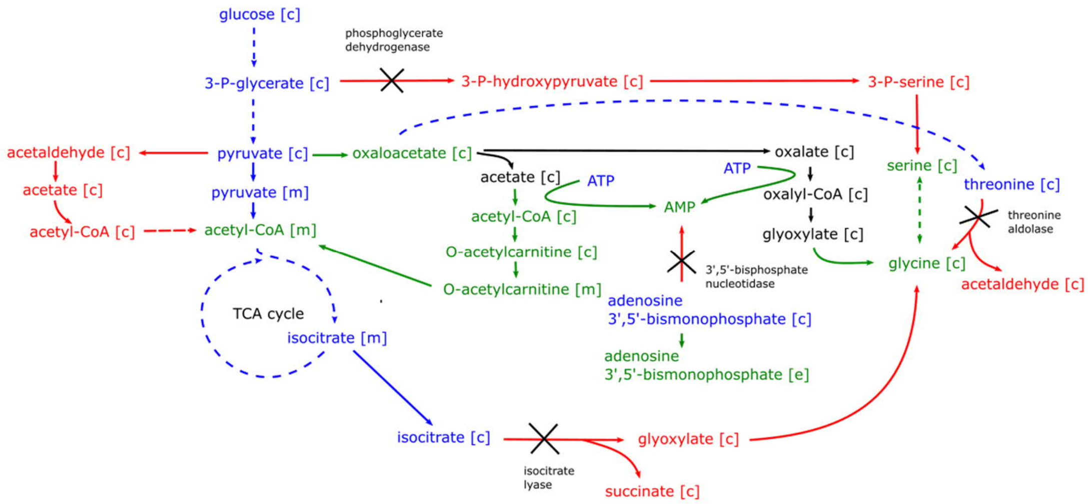

Figure 10. This solution has all the same reaction deletions as the intuitive solution to couple the first three reactions of oxalate pathway for glycolic acid synthesis (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) with growth, and in addition, 3',5'-bisphosphate nucleotidase is deleted. When 3',5'-bisphosphate nucleotidase is deleted adenosine 3'5'-bismonophosphate is secreted outside of the cell and the fluxes in acetyl-CoA synthetase and oxalyl-CoA ligase are increased. Blue, red and green colours refer to native *S. cerevisiae* metabolism and black reactions are heterologous oxalate pathway reactions. Red colour refers to interrupted reactions and green colour to upregulated reactions. Solid line refers to one reaction and dashed line to multiple reactions.

As EvolveXGA algorithm searches reactions to be deleted, it does not take into account that to delete a reaction the gene encoding for the enzyme catalyzing the reaction has to be deleted and the enzyme might have multiple functions (enzyme promiscuity). In addition, disruption of enzyme encoding genes might have other, non-metabolic functions, whose effects are not predictable with metabolic models. Therefore, as with all strain improvement algorithms, the *in silico* attained solutions should be evaluated for feasibility before realization in wet lab. Next the solutions A and B are discussed from this perspective.

4.2.2 Feasibility of solution A to be implemented in wet lab

In solution A two key enzymes (pyruvate dehydrogenase and pyruvate decarboxylase) taking part in converting pyruvate to acetyl-CoA are interrupted. Production of cytosolic acetyl-CoA is necessary for growth as it is a lipid (Flikweert *et al.*, 1999; Fakas, 2017) and lysine (Flikweert *et al.*, 1999) biosynthesis precursor. Mitochondrial acetyl-CoA contributes to the energy generation and amino acid synthesis. It enters the tricarboxylic acid cycle (TCA cycle) by forming citrate in a condensation reaction with oxaloacetate. Citrate is then oxidized and reduced electron carriers are formed. The carriers pass the electrons to the electron transfer chain that creates the proton gradient for ATP synthesis. In addition, mitochondrial acetyl-CoA is essential for glutamate biosynthesis as citrate is converted in the TCA cycle to 2-oxoglutarate, a TCA cycle intermediate, that is a glutamate biosynthesis precursor (DeLuna *et al.*, 2001).

Acetyl-CoA can be produced from pyruvate either directly by mitochondrial pyruvate dehydrogenase or through a cytosolic three reaction bypass including pyruvate decarboxylase, acetaldehyde dehydrogenase and acetyl-CoA synthetase. During completely respiratory growth pyruvate is mainly converted to acetyl-CoA by pyruvate dehydrogenase (Pronk *et al.*, 1994). However, in *S. cerevisiae* completely respiratory growth occurs only under glucose-limited conditions (Petrik *et* al., 1983; Urk *et al.*, 1988). With high glucose availability the respiratory capacity becomes limited and also fermentative metabolism occurs. In fermentative metabolism pyruvate is converted to acetaldehyde by pyruvate decarboxylase. NADH that has been formed under glycolysis can then be reoxidized with acetaldehyde as an electron acceptor (Pronk *et al.*, 1994).

There are three different genes encoding for pyruvate decarboxylase in *S. cerevisiae* (*PDC1*, *PDC5* and *PDC6*) (Hohmann, 1991). However, deletion of *PDC1* and *PDC5* is enough to produce strains without detectable levels of pyruvate decarboxylase activity at least in glucose, galactose or ethanol medium (Hohmann and Cederberg, 1990; Hohmann, 1991). Flikweert *et al*. (1999) presented that pyruvate decarboxylase could be deleted from *S. cerevisiae*, but acetate was required in the culture media as mitochondrial acetyl-CoA can not be transported to cytosol by carnitine shuttle in *S. cerevisiae*. However, acetate should not be necessary in the culture medium for *S. cerevisiae* with the heterologous oxalate pathway as it can be formed by oxaloacetase (Table 1) (Figure 9).

While *S. cerevisiae* contains three different pyruvate decarboxylase isozymes it does not have an isozyme for pyruvate dehydrogenase. However, pyruvate dehydrogenase is a large enzyme complex with five different subunits (E1α, E1β, E2, E3 and X) (Pronk *et al.*, 1996) and they are encoded by five different genes (Table 2). The E3 subunit is not specific only for pyruvate dehydrogenase, but is also part of other enzymes (2-oxoglutarate dehydrogenase, 2-oxoacid dehydrogenase and glycine carboxylase). However, it is enough to delete any of the genes *PDA1*, *PDB1*, *LAT1* or *PDX1* coding for E1α, E1β, E2 and X subunits respectively to remove pyruvate decarboxylase activity (Lawson *et al*., 1991a; Lawson *et al*., 1991b; Miran *et al*., 1993; Wenzel *et al.*, 1992). Therefore, the deletion of pyruvate dehydrogenase does not require disruption of any other enzymes. Pronk *et* al. (1994) showed that *S. cerevisiae* was able to grow without pyruvate dehydrogenase, but with a reduced biomass production yield as acetyl-CoA synthesis from acetate requires ATP.

According to my best knowledge no experiments have been made *in vivo* where both pyruvate dehydrogenase and pyruvate decarboxylase are deleted. In wild type *S. cerevisiae* deletion of these two enzymes would leave the production of acetaldehyde through threonine aldolase the only way to produce acetyl-CoA. Therefore, the deletion of both pyruvate decarboxylase and pyruvate dehydrogenase in a wild type *S. cerevisiae* might not be possible as it would not be possible to produce acetyl-CoA from threonine in high enough levels. However, growing *S. cerevisiae* with the heterologous oxalate pathway lacking pyruvate dehydrogenase and pyruvate decarboxylase in completely respiratory growth conditions could be possible. Respiratory growth conditions are required as cells without pyruvate decarboxylase are unable to reoxide NADH under fermentative growth. Therefore, glucose-limited media or nonfermentative sugars should be used as carbon source.

In addition to pyruvate decarboxylase and pyruvate dehydrogenase also mitochondrial acetyl-CoA synthetase and NAD^+ specific Isocitrate dehydrogenase are deleted in solution A. *S. cerevisiae* has two acetyl-CoA synthetases *ACS1* and *ACS2.* In the used *S. cerevisiae* GEM *ACS1* is annotated to the mitochondrial acetyl-CoA synthetase. However, it has been suggested that instead of mitochondria Acs1p would be located to nucleus and cytoplasm (Chen *et al*., 2012). Nevertheless, (as discussed in chapter 4.2.1) deletion of mitochondrial acetyl-CoA synthetase has almost no effect on the fitness score, and therefore, its deletion is not necessary.

NAD^+ specific Isocitrate dehydrogenase contains two distinct polypeptide subunits and these subunits are coded by *IDH1* and *IDH2*. It is enough to delete either of these genes to remove NAD^+ specific isocitrate dehydrogenase activity (Keys and McAlister-Henn, 1990) and these genes do not code for any other enzymes. Deletion of *IDH2* did not reduce the growth rate in rich medium with glucose as carbon source, but growth with non-fermentable carbon sources was reduced compared to wild type growth rate (Haselbeck and McAlister-Henn, 1993). Therefore, it could be possible to realize solution A *in vivo*, without deleting mitochondrial acetyl-CoA synthetase, if glucose-limited conditions or media with nonfermentable sugars as carbon source is used. For such scenario reduced growth rate would be expected. In solution B the coupling of the target reactions with growth is not as strong as in solution A, but it does not include deletions of reactions taking part in acetyl-CoA metabolism or TCA cycle reactions, and therefore, its realization *in vivo* could have less growth effects.

4.2.3 Feasibility of Solution B to be implemented in wet lab

In solution B (Figure 10) and in the intuitive solution (Figure 8) *ICL1, SER3, SER33* and *GLY1* are deleted. In addition, in solution B *MET22* is interrupted (Table 2). *ICL1* encodes isocitrate lyase. Isocitrate lyase activity is essential when ethanol or acetate is used as the only carbon source, but cells lacking *ICL1* can grow on glucose (Fernández *et al*., 1992; Luttik *et al*., 2000). *MET22* is a non-essential gene, but mutants lacking it are organic sulfur auxotrophs and therefore either homocysteine,

cysteine, methionine or S-adenosylmethionine is needed in the culture media to sustain growth. In addition, deletion of *MET22* causes salt sensitivity (Thomas and Surdin-Kerjan, 1997).

Deletion of *SER3* and *SER33* has been used to improve succinic acid production (Otero *et al*., 2013). *SER3* and *SER33* encode phosphoglycerate dehydrogenase that forms 3-phosphonato-D-glycerate, but in addition they have also a side activity of forming 2-hydroxyglutarate from 2-oxoglutarate (Becker-Kettern *et al*., 2016). The interruption of this side activity would be minimally beneficial for coupling oxalate pathway fluxes with growth (EvolveX score increased by 0.004 mmol/g CDW when reaction NAD^+ + 2-hydroxyglutarate <=> 2-oxoglutarate + H^+ + NADH was deleted in addition to other solution B reaction deletions). This side activity could be replaced in cells by hydroxyglutarate transhydrogenase, encoded by *DLD3*. According to Becker-Kettern *et al*. (2016) 2-hydroxyglutarate transhydrogenase catalyze the conversion of 2-hydroxyglutarate to 2-oxoglutarate, but it is possible, that it can also catalyzes the reversible reaction. Therefore, a small improvement in the growth-coupling could be achieved if *DLD3* was deleted in addition to the other solution B deletions.

GLY1 encodes enzymes catalyzing the production of glycine and acetaldehyde both from threonine and allo-threonine (Liu J-Q *et al.*, 1997). However, when possible flux ranges of all the fluxes with solution B reaction deletions were simulated with FVA, there was no flux in the reaction using allo-threonine (reaction id: r_0675) even when it is not deleted. McNeil *et al*. (1994) showed that *S. cerevisiae* lacking *GLY1* as well as *SHM1* and *SHM2* (i.e. genes encoding serine hydroxymethyltransferases needed for the conversion of serine to glycine) could grow as long as glycine was added to the culture medium. Therefore, deletion of *GLY1* together with *SER3* and *SER33* deletion should be possible *in vivo* when glycine can be produced through the oxalate pathway.

4.2.4 Sub-optimal EvolveXGA reaction deletion solutions

In addition to the final solutions also the best solution of each generation was collected. All the best solutions had pyruvate dehydrogenase and pyruvate decarboxylase deleted. However, good scores without deleting these genes were also attained. Many of these solutions had also a higher score than solution B. One of these solutions (solution C) is presented in Table 3. Deletion of acetaldehyde dehydrogenaseses makes the oxalate pathway the only possibility to produce cytosolic acetyl-CoA. Deletion of citrate synthase blocks acetyl-CoA oxidation in the TCA cycle and NADH is formed in β-oxidation instead.

Table 3. A sub-optimal solution to couple three first oxalate pathway reactions for glycolic acid synthesis (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) with growth obtained with EvolveXGA. This solution is not a final solution but it was attained as one of the best solutions of intermediate generations.

Solution C				
EvolveX score	8.58 mmol/g CDW			
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0025	2-isopropylmalate synthase	3-methyl-2-oxobutanoate [mitochondrion] + acetyl-CoA [mitochondrion] + H2O [mitochondrion] --> 2-isopropylmalate [mitochondrion] + coenzyme A [mitochondrion] + H+ [mitochondrion]	*LEU4*, *LEU9*
	r_0173	aldehyde dehydrogenase (NADP)	acetaldehyde [cytoplasm] + H2O [cytoplasm] + NADP(+) [cytoplasm] --> acetate [cytoplasm] + 2.0 H+ [cytoplasm] + NADPH [cytoplasm]	*ALD6*
	r_0300	citrate synthase	acetyl-CoA [mitochondrion] + H2O [mitochondrion] + oxaloacetate [mitochondrion] --> citrate [mitochondrion] + coenzyme A [mitochondrion] + H+ [mitochondrion]	*CIT1*, *CIT3*
	r_2116	aldehyde dehydrogenase (NAD)	acetaldehyde [cytoplasm] + H2O [cytoplasm] + NAD [cytoplasm] --> acetate [cytoplasm] + H+ [cytoplasm] + NADH [cytoplasm]	*ALD2*, *ALD3*

Realization of solution C could be possible *in vivo*. There are two mitochondrial citrate synthases Cit1p and Cit3p and level of *CIT3* expression has been shown to be much lower than *CIT1* expression (Graybill e*t al*., 2007). Kispal e*t al*. (1988) showed that deletion of *CIT1* did not affect the growth when galactose, glycerol, lactate, pyruvate or glutamate were used as carbon sources. However, it is possible that deletion of both *CIT1* and *CIT3* would have fitness effects. Ohta *et al.* (2016) deleted *LEU4* and *LEU9* successfully together with several other genes to improve ethanol stress tolerance in S*. cerevisiae*. Also the deletion of the three cytosolic aldehyde dehydrogenases (i.e. encoded by *ALD2*, *ALD3* and *ALD6* (Navarro-Aviño *et al*., 1999)) should not have detrimental effects as acetate is formed through the oxalate pathway.

4.3 Combined reaction deletion and selection niche strategies for improving glycolic acid production

Next, combined strategies of reaction deletions and selection niche for exposing the three first reactions of oxalate pathway (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) (Table 1) to optimal selection during ALE was searched in EvolveXGA optimization runs of 3000 generations. In each generation of EvolveXGA, the fitness of individuals was scored using EvolveX (Jouhten and Patil, 2016; Jouhten *et al*., 2021 Preprint) as target fluxes relative to the arbitrarily chosen growth yield. To ensure numerical accuracy of the fluxes contributing to the minor compounds included in the biomass equation of *S. cerevisiae* in the model, the growth yield was initially set to ten. However, with growth yield of ten, the artificial flux bounds -1000 and 1000 used in the model simulations were not low/high enough for searching combined reaction deletion and selection niche strategies as they formed an unintended and biologically not meaningful restriction to the solutions. Since increasing the absolute values of flux lower and upper bounds by ten fold did not resolve the restriction, arbitrary growth yield was finally set to one and the bounds to -20000 and 20000 and the unintended restrictions were removed. Even when the arbitrary growth yield was divided by ten, all the fluxes in biomass pseudoreactions were high enough (> 10^{-9}) to be numerically accurate.

4.3.1 EvolveXGA solutions of reaction deletions and selection niche

The EvolveXGA algorithm was run six times and the attained results are presented in Table 4. Solutions 2 and 3 were both obtained two times.

Table 4. Solutions to couple three first oxalate pathway reactions for glycolic acid synthesis (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) with growth with suitable nutrients and reaction deletions.

Solution 1				
EvolveX score	5060.78 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1902	methionine		
	r_2033	pyruvate		
	r_2051	spermidine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0091	6-phosphogluconolactonase	6-O-phosphono-D-glucono-1,5-lactone [cytoplasm] + H2O [cytoplasm] --> 6-phospho-D-gluconate [cytoplasm] + H+ [cytoplasm]	*SOL3*, *SOL4*
	r_0447	formate-tetrahydrofolate ligase	ATP [mitochondrion] + formate [mitochondrion] + THF [mitochondrion] <=> 10-formyl-THF [mitochondrion] + ADP [mitochondrion] + phosphate [mitochondrion]	*MIS1*
	r_0959	pyruvate decarboxylase	H+ [cytoplasm] + pyruvate [cytoplasm] --> acetaldehyde [cytoplasm] + carbon dioxide [cytoplasm]	*PDC1*, *PDC5*, *PDC6*
	r_0961	pyruvate dehydrogenase	coenzyme A [mitochondrion] + NAD [mitochondrion] + pyruvate [mitochondrion] --> acetyl-CoA [mitochondrion] + carbon dioxide [mitochondrion] + NADH [mitochondrion]	*PDB1*, *PDA1*, *LPD1*, *PDX1*, *LAT1*

Solution 2				
EvolveX score	5060.78 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1902	methionine		
	r_2033	pyruvate		
	r_2051	spermidine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0466	glucose 6-phosphate dehydrogenase	D-glucose 6-phosphate [cytoplasm] + NADP(+) [cytoplasm] --> 6-O-phosphono-D-glucono-1,5-lactone [cytoplasm] + H+ [cytoplasm] + NADPH [cytoplasm]	*ZWF1*
	r_0724	methenyltetrahydrofolate cyclohydrolase	5,10-methenyl-THF [mitochondrion] + H2O [mitochondrion] <=> 10-formyl-THF [mitochondrion] + H+ [mitochondrion]	*MIS1*
	r_0959	pyruvate decarboxylase	H+ [cytoplasm] + pyruvate [cytoplasm] --> acetaldehyde [cytoplasm] + carbon dioxide [cytoplasm]	*PDC1*, *PDC5*, *PDC6*
	r_0961	pyruvate dehydrogenase	coenzyme A [mitochondrion] + NAD [mitochondrion] + pyruvate [mitochondrion] --> acetyl-CoA [mitochondrion] + carbon dioxide [mitochondrion] + NADH [mitochondrion]	*PDB1*, *PDA1*, *LPD1*, *PDX1*, *LAT1*

Solution 3				
EvolveX score	5060.78 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1902	methionine		
	r_2033	pyruvate		
	r_2051	spermidine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0091	6-phosphogluconolactonase	6-O-phosphono-D-glucono-1,5-lactone [cytoplasm] + H2O [cytoplasm] --> 6-phospho-D-gluconate [cytoplasm] + H+ [cytoplasm]	*SOL3*, *SOL4*
	r_0733	methylenetetrahydrofolate dehydrogenase (NADP)	5,10-methylenetetrahydrofolate [mitochondrion] + NADP(+) [mitochondrion] <=> 5,10-methenyl-THF [mitochondrion] + NADPH [mitochondrion]	*MIS1*
	r_0959	pyruvate decarboxylase	H+ [cytoplasm] + pyruvate [cytoplasm] --> acetaldehyde [cytoplasm] + carbon dioxide [cytoplasm]	*PDC1*, *PDC5*, *PDC6*
	r_0961	pyruvate dehydrogenase	coenzyme A [mitochondrion] + NAD [mitochondrion] + pyruvate [mitochondrion] --> acetyl-CoA [mitochondrion] + carbon dioxide [mitochondrion] + NADH [mitochondrion]	*PDB1*, *PDA1*, *LPD1*, *PDX1*, *LAT1*

Solution 4				
EvolveX score	22.55 mmol/g CDW			
Selection niche	reaction id	compound name		
	r_1710	D-galactose		
	r_1902	methionine		
	r_2051	spermidine		
Deleted reactions	reaction id	reaction name	reaction	gene(s) encoding for the enzyme(s) catalyzing the reaction
	r_0117	aconitase	2-methylcitrate [mitochondrion] <=> (2S,3R)-3-hydroxybutane-1,2,3-tricarboxylic acid [mitochondrion]	*PDH1*
	r_0446	formate-tetrahydrofolate ligase	ATP [cytoplasm] + formate [cytoplasm] + THF [cytoplasm] <=> 10-formyl-THF [cytoplasm] + ADP [cytoplasm] + phosphate [cytoplasm]	*ADE3*
	r_0662	isocitrate lyase	isocitrate [cytoplasm] --> glyoxylate [cytoplasm] + succinate [cytoplasm]	*ICL1*
	r_0891	phosphoglycerate dehydrogenase	3-phosphonato-D-glycerate(3-) [cytoplasm] + NAD [cytoplasm] --> 3-phospho-hydroxypyruvate [cytoplasm] + H+ [cytoplasm] + NADH [cytoplasm]	*SER3, SER33*

Solution 1, 2 and 3 with pyruvate, methionine, and spermidine as the selection niche redirect the flux equally as glucose 6-phosphate dehydrogenase and 6-phosphogluconolactonase are consecutive reactions of the linear oxidative pentose phosphate pathway (Figure 11) and formate-tetrahydrofolate ligase, methenyltetrahydrofolate cyclohydrolase and methylenetetrahydrofolate dehydrogenase ($NADP^+$) are all part of tetrahydrofolate biosynthesis.

The reason why deleting reactions from the oxidative pentose phosphate pathway (i.e. glucose-6-phosphate dehydrogenase or 6-phosphogluconolactonase) and tetrahydrofolate biosynthesis (i.e. formate-tetrahydrofolate ligase, methenyltetrahydrofolate cyclohydrolase or methylenetetrahydrofolate dehydrogenase) is beneficial to oxalate pathway seems to be because both of these pathways form NADPH. NADPH is needed as reducing power for biomass synthesis. Therefore, without the pentose phosphate pathway or tetrahydrofolate biosynthesis more NADPH has to be produced through $NADP^+$ dependent isocitrate dehydrogenase which pulls flux to the TCA cycle. Thus, more acetyl-CoA has to be produced and the only way to produce it is through the oxalate pathway when pyruvate dehydrogenase and pyruvate decarboxylase are deleted. A small part of the pyruvate is also consumed by gluconeogenesis (Figure 11). Methionine is used as nitrogen source. In addition, methionine is needed from the culture medium as its synthesis is interrupted when tetrahydrofolate biosynthesis reactions are deleted (Thomas and Surdin-Kerjan, 1997).

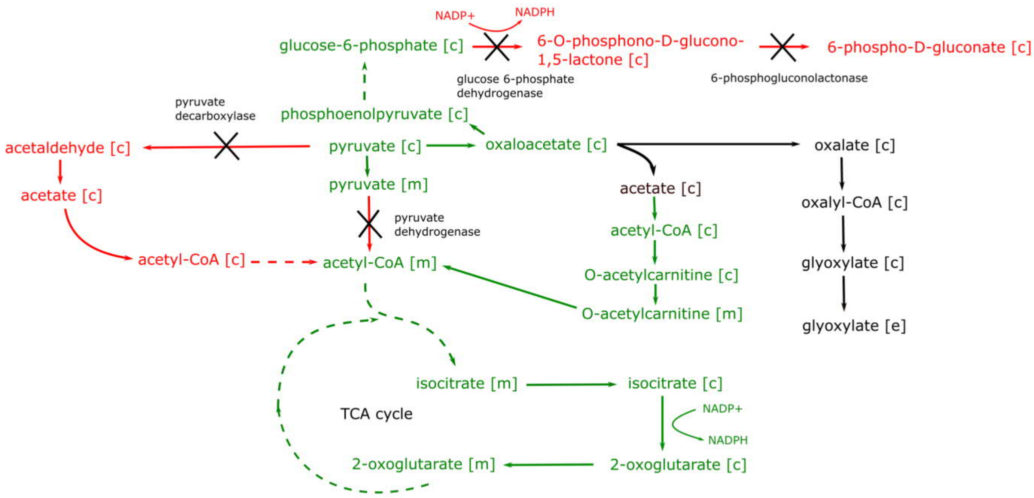

Figure 11. The solution with highest EvolveX score to couple three first oxalate pathway reactions for glycolic acid synthesis (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) with growth with suitable nutrients and reaction deletions. NADPH is produced through $NADP^+$ dependent isocitrate dehydrogenase and therefore flux in the TCA cycle is increased. Deletion of pyruvate dehydrogenase and pyruvate decarboxylase forces acetyl-CoA to be produced through the first reaction of oxalate pathway. Either glucose-6-phosphate dehydrogenase or 6-phosphogluconolactonase is deleted in the solutions. Reactions of tetrahydrofolate biosynthesis and methionine catabolism is omitted from the picture. Red and green colour refer to native *S. cerevisiae* metabolism and black reactions are heterologous oxalate pathway reactions. Red colour refers to interrupted reactions and green colour to upregulated reactions. Solid line refers to one reaction and dashed line to multiple reactions.

Even if the algorithm suggests spermidine to be one of selection niche components in addition to pyruvate and methionine, there is no flux in spermidine uptake. The reason why the algorithm, however, always finds spermidine uptake together with the other nutrients and deletions of solutions 1, 2 and 3 must therefore be that any other third nutrient in the selection niche would affect how methionine and pyruvate is used and three components are required in the optimization. In these solutions pyruvate uptake is very high (i.e. pyruvate uptake is 1714 mmol/g CDW and methionine uptake 19 mmol/g CDW when growth is 1 h^{-1}). Very high nutrient uptake compared with growth means that *in vivo* the solutions would likely be inviable or uninteresting as the growth would be very slow. Therefore, EvolveXGA was next run three times while constraining the total nutrient uptake to reasonable 75 mmol/g CDW when the arbitrary growth yield was set to 1 h^{-1}.

4.3.2 EvolveXGA solutions with total nutrient uptake restricted

The solutions obtained with EvolveXGA when the total nutrient uptake constrain had been introduced are presented in Table 5. Solution 5 was obtained once and solution 6 two times.

Table 5. Two solutions to couple three first oxalate pathway reactions for glycolic acid synthesis (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) with growth with suitable nutrients and reaction deletions when the sum of nutrient uptake was restricted to 75 mmol/g CDW and growth was 1 h^{-1}.

Solution 5				
EvolveX score	415.85 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1634	acetate		
	r_1883	cysteine		
	r_1993	palmitate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0253	carnitine O-acetyltransferase	(R)-carnitine [peroxisome] + acetyl-CoA [peroxisome] --> coenzyme A [peroxisome] + O-acetylcarnitine [peroxisome]	*CAT2*
	r_0256	catalase	2.0 hydrogen peroxide [peroxisome] --> 2.0 H2O [peroxisome] + oxygen [peroxisome]	*CTA1*
	r_0301	citrate synthase, peroxisomal	acetyl-CoA [peroxisome] + H2O [peroxisome] + oxaloacetate [peroxisome] --> citrate [peroxisome] + coenzyme A [peroxisome] + H+ [peroxisome]	*CIT2*
	r_0662	isocitrate lyase	isocitrate [cytoplasm] --> glyoxylate [cytoplasm] + succinate [cytoplasm]	*ICL1*

Solution 6				
EvolveX score	22.55 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1710	D-galactose		
	r_1902	methionine		
	r_1993	palmitate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0446	formate-tetrahydrofolate ligase	ATP [cytoplasm] + formate [cytoplasm] + THF [cytoplasm] <=> 10-formyl-THF [cytoplasm] + ADP [cytoplasm] + phosphate [cytoplasm]	*ADE3*
	r_0451	fumarase	fumarate [mitochondrion] + H2O [mitochondrion] <=> (S)-malate [mitochondrion]	*FUM1*
	r_0662	isocitrate lyase	isocitrate [cytoplasm] --> glyoxylate [cytoplasm] + succinate [cytoplasm]	*ICL1*
	r_0917	phosphoserine phosphatase (L-serine)	3-phospho-serine [cytoplasm] + H2O [cytoplasm] --> L-serine [cytoplasm] + phosphate [cytoplasm]	*SER1*

Solution 5 couples the oxalate pathway reactions with growth by reaction deletions allowing the cells to use palmitate only when producing glyoxylate and by restricting other possibilities to produce glyoxylate than oxalate pathway (Figure 12). In addition, deletion of catalase forces hydrogenperoxide to be reduced with thioredoxin pathway which requires NADPH. The required NADPH can be formed in the TCA cycle which also allows the use of acetyl-CoA that has been formed as a byproduct of oxalate pathway (Figure 12). Even if acetate is one of the selection niche nutrients there is no flux in acetate uptake and therefore cysteine and palmitate could be used as sole selection niche nutrients.

However, it would not be possible to realize solution 5 *in vivo*. *CAT2* (Table 5) encodes both the peroxisomal and mitochondrial carnitine O-acetyltransferase (Elgersma *et al.*, 1995). Therefore, if *CAT2* was deleted the cell would not be able to transport acetyl-CoA from cytosol to mitochondrion. The cell could probably produce mitochondrial acetyl-CoA for the TCA cycle through pyruvate dehydrogenase, but there would not be a pathway consuming cytosolic acetyl-CoA in high enough quantities.

Solution 6 is almost the same as solution 4 (Table 4). Even if the scores of these solutions are low compared to the previously discussed solutions combining the reaction deletions and selection niche it is higher than the scores of the solutions where only reaction deletions were searched. Thus, solution 4 and 6 could potentially be used in coupling the oxalate pathway reactions with growth. However, a closer evaluation about the effects on growth of the required reaction deletions should be made.

In conclusion, EvolveXGA was able to find several strategies to couple three first reactions of oxalate pathway (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate) with growth both by only defining reaction deletions, with glucose and ammonium as a fixed selection niche, and by defining combined reaction deletions and selection niche. The best strategy to be implemented in wet lab should then be chosen after evaluating the realizability of the required reaction deletions.

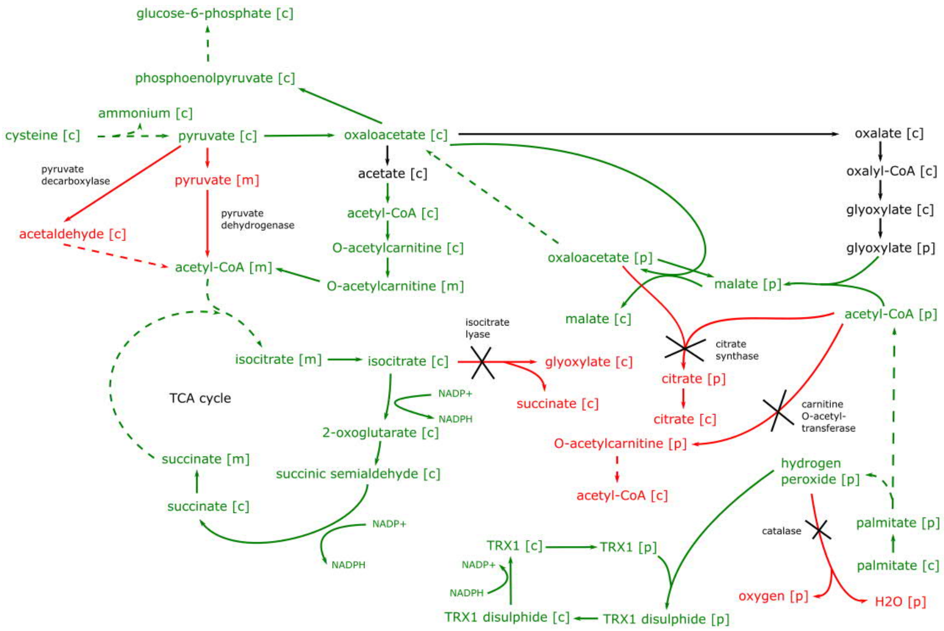

Figure 12. The solution with highest EvolveX score to couple three first oxalate pathway reactions for glycolic acid synthesis (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) with growth with suitable nutrients and reaction deletions when nutrient uptake was restricted to 75 mmol/g CDW and growth was 1 h^{-1}. Deletion of peroxisomal citrate synthase and peroxisomal carnitine O-acetyltransferase makes converting acetyl-CoA and glyoxylate to malate the only way to transfer the carbon atoms from palmitate to cytosol. Therefore, glyoxylate has to be produced. When isocitrate lyase is deleted the oxalate pathway becomes the only way to produce glyoxylate. Deletion of calatase forces hydrogen peroxide to be reduced with thioredoxin (TRX1) as an electron acceptor and TRX1 disulphide is formed. Re-reducing TRX1 disulphide to TRX1 requires NADPH. This NADPH is formed in the TCA cycle. Flux in the TCA cycle allows the use of acetyl-CoA that has been formed as a byproduct in oxalate pathway. L-cysteine is used as a nitrogen and carbon nutrient. Red and green colours refer to native *S. cerevisiae* metabolism and black reactions are heterologous oxalate pathway reactions. Red colour refers to interrupted reactions and green colour to upregulated reactions. Solid line refers to one reaction and dashed line to multiple reactions.

4.4 Combined reaction deletion and selection niche strategies for improving production of other compounds

Next, the generalizability of the EvolveXGA method to different products was evaluated. ALE strategies to improve the production of a variety of 28 other heterologous compounds (Figure 13) (APPENDIX 2) were searched using EvolveXGA. These heterologous compounds are drug precursors, flavor aroma compounds, biofuels and biopolymer precursors from different sides of central metabolism (Jouhten *et al.*, 2016).

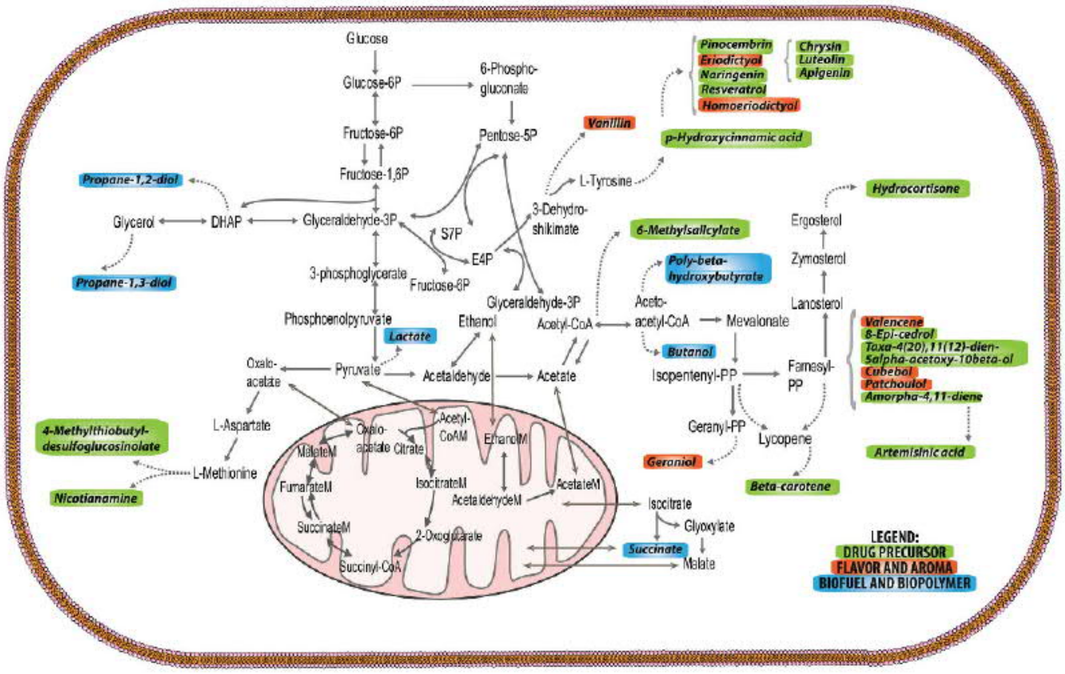

Figure 13. 28 heterologous products for which EvolveXGA was tested and their connection with the native *S. cerevisiae* metabolism. Glycolic acid is not present in this figure and succinate was not part of the tested compounds. Figure from Jouhten e*t al*. (2016).

4.4.1 EvolveXGA solutions coupling the production pathway reactions with growth

For identifying production improving ALE strategies using EvolveXGA for each of the compounds the reactions of the corresponding heterologous pathway were set as targets. A generation number of 5000 was used as the termination condition for the solution search. EvolveXGA found a solution to couple 12 of these tested heterologous production pathways with growth using reaction deletions and selection niche. No restriction on the sum of nutrient uptake was set, but the sum of the nutrient uptakes of all the solutions were under 75 mmol/g CDW.

Eight of the attained solutions (i.e. apigenin-, chrysin-, eriodictyol-, homoeriodictyol-, luteolin-, naringenin-, pinocembrin- and resveratrol pathway) had glycerol, tyrosine and palmitate as selection niche (Table 6). In addition, p-hydroxycinnamic acid pathway had tyrosine and palmitate in its selection niche. All these nine pathways start with the same tyrosine ammonia lyase reaction in cytosol (tyrosine <=> p-hydroxycinnamic acid + ammonium) (APPENDIX 2). Therefore, tyrosine is a direct precursor for these heterologous pathways. To understand better the growth coupling in these pathways the effect of the selection niche and reaction deletions were studied in more detail for naringenin pathway (Figure 14).

In the solution to couple heterologous naringenin pathway reactions with growth the coupling was mainly created with selection niche nutrients. The EvolveX score with glycerol, tyrosine and palmitate as nutrients without any reaction deletions was 37.54 mmol/g CDW, when the score with reaction deletions was 39.52 mmol/g CDW. Specifically, the growth coupling was formed with tyrosine as a nitrogen source and glycerol and/or palmitate as a carbon source. Glycerol and palmitate had equal effect on the EvolveX score. Tyrosine in the selection niche created growth coupling because the used *S. cerevisiae* GEM lacked the Ehrlich pathway reactions for tyrosine catabolism (Hazelwood *et al*., 2008). Thus, with this model the only way to use tyrosine as a nitrogen source was by using the tyrosine ammonia lyase reaction (Figure 14). Therefore, if this solution was realized *in vivo,* the genes encoding for enzymes catalyzing the first reaction step of Ehrlich pathway (i.e. *ARO8* and *ARO9*) (Figure 14) (Hazelwood *et al*., 2008) would need to be deleted.

In addition to the selection niche, also the deletion of 3’,5’-bisphosphate nucleotidase had a positive effect on the score. The score with glycerol, tyrosine and palmitate as nutrients and with 3’,5’-bisphosphate nucleotidase deleted was only 0.01 mmol/g CDW smaller than the score with all the reaction deletions of the solution. Therefore, other deletions did not have any notable effect on the score. 3’,5’-bisphosphate nucleotidase was also deleted in all the pathways starting with tyrosine ammonia lyase. In all of these pathways, except in p-hydroxycinnamic acid pathway, there is an AMP producing reaction in the heterologous pathway (APPENDIX 2). Therefore, when 3’,5’-bisphosphate nucleotidase is deleted AMP that is produced through the heterologous pathways can be used in RNA biosynthesis in the same way as in Solution B with oxalate pathway (Figure 10).

However, as discussed earlier in chapter 4.2.3 deletion of *MET22*, encoding 3’,5’-bisphosphate nucleotidase causes organic sulfur auxotrophy (Thomas and Surdin-Kerjan, 1997). Therefore, sulfur containing amino acid is needed in the culture medium. The effect of adding methionine, cysteine or S-adenosyl-L-methionine to the selection niche was tested for naringenin pathway and all of these additions had a negative effect on the EvolveX score. The score was 37.03 mmol/g CDW when either methionine or cysteine was added and dropped to zero when S-adenosyl-L-methionine was added. Thus, the best solution to couple naringenin pathway reactions with growth would be to use only the selection niche (with tyrosine and palmitate or glycerol) without any reaction deletions (apart from *ARO8* and *ARO9*) .

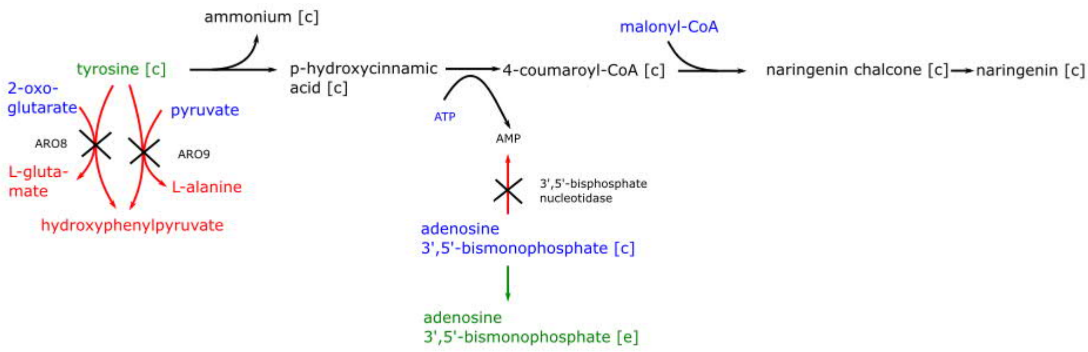

Figure 14. Solution to couple naringenin pathway reactions with growth. The only possibility to use tyrosine as a nitrogen source is to use tyrosine ammonia lyase when *ARO8* and *ARO9* are deleted. In addition, deletion of 3',5'-bisphosphate nucleotidase increases the growth coupling as the AMP produced by naringenin pathway can be used in RNA biosynthesis. Blue, red and green colours refer to native *S. cerevisiae* metabolism and black colour to heterologous naringenin pathway reactions. Red colour refers to interrupted reactions and green colour to upregulated reactions.

Table 6. Solutions to couple heterologous production pathways with growth for 12 different heterologous pathways.

4-Methyltiobutyl-desulfoglucosinolate pathway				
EvolveX score	131,52 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1761	ethanol		
	r_1902	methionine		
	r_1993	palmitate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0042	3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase	D-erythrose 4-phosphate [cytoplasm] + H2O [cytoplasm] + phosphoenolpyruvate [cytoplasm] --> 7-phospho-2-dehydro-3-deoxy-D-arabino-heptonic acid [cytoplasm] + phosphate [cytoplasm]	*ARO3*, *ARO4*
	r_0112	acetyl-CoA synthetase	acetate [cytoplasm] + ATP [cytoplasm] + coenzyme A [cytoplasm] --> acetyl-CoA [cytoplasm] + AMP [cytoplasm] + diphosphate [cytoplasm]	*ACS1*, *ACS2*
	r_4235	ATP:D-glucose 6-phosphotransferase	ATP [cytoplasm] + D-glucose [cytoplasm] --> ADP [cytoplasm] + D-glucose 6-phosphate [cytoplasm]	*EMI2*
	r_4484	O4-succinyl-L-homoserine:hydrogen sulfide S-(3-amino-3-carboxypropyl)transferase; O-succinyl-L-homoserine succinate-lyase (adding hydrogen sulfide)	hydrogen sulfide [cytoplasm] + O-succinyl-L-homoserine [cytoplasm] <=> L-homocysteine [cytoplasm] + succinate [cytoplasm]	*STR2*, *YLL058W*, *YML082W*

Apigenin pathway

EvolveX score	46,1 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1808	glycerol		
	r_1913	tyrosine		
	r_1993	palmitate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0032	3',5'-bisphosphate nucleotidase	adenosine 3',5'-bismonophosphate [cytoplasm] + H2O [cytoplasm] --> AMP [cytoplasm] + phosphate [cytoplasm]	*MET22*
	r_0366	enolase	2-phospho-D-glyceric acid [cytoplasm] <=> H2O [cytoplasm] + phosphoenolpyruvate [cytoplasm]	*ERR1, ERR2, ERR3, ENO1, ENO2*
	r_0446	formate-tetrahydrofolate ligase	ATP [cytoplasm] + formate [cytoplasm] + THF [cytoplasm] <=> 10-formyl-THF [cytoplasm] + ADP [cytoplasm] + phosphate [cytoplasm]	*ADE3*
	r_1623	5-formethyltetrahydrofolate cyclo-ligase	5-formyltetrahydrofolic acid [mitochondrion] + ATP [mitochondrion] --> 5,10-methenyl-THF [mitochondrion] + ADP [mitochondrion] + phosphate [mitochondrion]	*FAU1*

Chrysin pathway

EvolveX score	53,92 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1808	glycerol		
	r_1913	tyrosine		
	r_1993	palmitate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0032	3',5'-bisphosphate nucleotidase	adenosine 3',5'-bismonophosphate [cytoplasm] + H2O [cytoplasm] --> AMP [cytoplasm] + phosphate [cytoplasm]	*MET22*
	r_0366	enolase	2-phospho-D-glyceric acid [cytoplasm] <=> H2O [cytoplasm] + phosphoenolpyruvate [cytoplasm]	*ERR1, ERR2, ERR3, ENO1, ENO2*
	r_0851	phenylalanine transaminase	2-oxoglutarate [cytoplasm] + L-phenylalanine [cytoplasm] <=> keto-phenylpyruvate [cytoplasm] + L-glutamate [cytoplasm]	*ARO8*
	r_2117	phenylalanine transaminase	L-phenylalanine [cytoplasm] + pyruvate [cytoplasm] <=> keto-phenylpyruvate [cytoplasm] + L-alanine [cytoplasm]	*ARO9*

Eriodictyol pathway

EvolveX score	46,1 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1808	glycerol		
	r_1913	tyrosine		
	r_1993	palmitate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0032	3',5'-bisphosphate nucleotidase	adenosine 3',5'-bismonophosphate [cytoplasm] + H2O [cytoplasm] --> AMP [cytoplasm] + phosphate [cytoplasm]	*MET22*
	r_0471	glutamate dehydrogenase (NADP)	2-oxoglutarate [cytoplasm] + ammonium [cytoplasm] + H+ [cytoplasm] + NADPH [cytoplasm] --> H2O [cytoplasm] + L-glutamate [cytoplasm] + NADP(+) [cytoplasm]	*GDH1, GDH3*
	r_0893	phosphoglycerate mutase	3-phosphonato-D-glycerate(3-) [cytoplasm] <=> 2-phospho-D-glyceric acid [cytoplasm]	*GPM1, YOR283W*
	r_1084	UTP-glucose-1-phosphate uridylyltransferase	D-glucose 1-phosphate [cytoplasm] + H+ [cytoplasm] + UTP [cytoplasm] <=> diphosphate [cytoplasm] + UDP-D-glucose [cytoplasm]	*UGP1, YHL012W*

Homoeriodictyol pathway

EvolveX score	52,69 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1808	glycerol		
	r_1913	tyrosine		
	r_1993	palmitate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0032	3',5'-bisphosphate nucleotidase	adenosine 3',5'-bismonophosphate [cytoplasm] + H2O [cytoplasm] --> AMP [cytoplasm] + phosphate [cytoplasm]	*MET22*
	r_0567	indole-3-pyruvate decarboxylase	H+ [cytoplasm] + indole-3-pyruvate [cytoplasm] --> carbon dioxide [cytoplasm] + indol-3-ylacetaldehyde [cytoplasm]	*PDC1, PDC5, PDC6*
	r_0819	ornithine transaminase	2-oxoglutarate [cytoplasm] + ornithine [cytoplasm] --> L-glutamate [cytoplasm] + L-glutamic 5-semialdehyde [cytoplasm]	*CAR2*
	r_0892	phosphoglycerate kinase	1,3-bisphospho-D-glycerate [cytoplasm] + ADP [cytoplasm] <=> 3-phosphonato-D-glycerate(3-) [cytoplasm] + ATP [cytoplasm]	*PGK1*

Lactate pathway

EvolveX score	48,66 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1683	choline		
	r_1714	D-glucose		
	r_1902	methionine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0451	fumarase	fumarate [mitochondrion] + H2O [mitochondrion] <=> (S)-malate [mitochondrion]	*FUM1*
	r_0452	fumarase, cytoplasmic	fumarate [cytoplasm] + H2O [cytoplasm] <=> (S)-malate [cytoplasm]	*FUM1*
	r_0467	glucose-6-phosphate isomerase	D-glucose 6-phosphate [cytoplasm] <=> D-fructose 6-phosphate [cytoplasm]	*PGI1*
	r_0486	glyceraldehyde-3-phosphate dehydrogenase	glyceraldehyde 3-phosphate [cytoplasm] + NAD [cytoplasm] + phosphate [cytoplasm] <=> 1,3-bisphospho-D-glycerate [cytoplasm] + H+ [cytoplasm] + NADH [cytoplasm]	*TDH1*, *TDH2*, *TDH3*

Luteolin pathway

EvolveX score	52,69 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1808	glycerol		
	r_1913	tyrosine		
	r_1993	palmitate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0032	3',5'-bisphosphate nucleotidase	adenosine 3',5'-bismonophosphate [cytoplasm] + H2O [cytoplasm] --> AMP [cytoplasm] + phosphate [cytoplasm]	*MET22*
	r_0348	dihydroneopterin monophosphate dephosphorylase	7,8-dihydroneopterin 3'-phosphate [cytoplasm] + H2O [cytoplasm] --> 7,8-dihydroneopterin [cytoplasm] + phosphate [cytoplasm]	*GET3*
	r_0491	glycerol-3-phosphate dehydrogenase (NAD)	dihydroxyacetone phosphate [cytoplasm] + H+ [cytoplasm] + NADH [cytoplasm] --> glycerol 3-phosphate [cytoplasm] + NAD [cytoplasm]	*GPD1*, *GPD2*
	r_0893	phosphoglycerate mutase	3-phosphonato-D-glycerate(3-) [cytoplasm] <=> 2-phospho-D-glyceric acid [cytoplasm]	*YOR283W*, *GPM1*

Naringenin pathway

EvolveX score	39,52 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1808	glycerol		
	r_1913	tyrosine		
	r_1993	palmitate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0032	3',5'-bisphosphate nucleotidase	adenosine 3',5'-bismonophosphate [cytoplasm] + H2O [cytoplasm] --> AMP [cytoplasm] + phosphate [cytoplasm]	*MET22*
	r_0162	alcohol acetyltransferase (phenylethanol alcohol)	2-phenylethanol [cytoplasm] + acetyl-CoA [cytoplasm] --> coenzyme A [cytoplasm] + phenethyl acetate [cytoplasm]	*ATF1*, *ATF2*
	r_0819	ornithine transaminase	2-oxoglutarate [cytoplasm] + ornithine [cytoplasm] --> L-glutamate [cytoplasm] + L-glutamic 5-semialdehyde [cytoplasm]	*CAR2*
	r_0884	phosphoenolpyruvate carboxykinase	ATP [cytoplasm] + oxaloacetate [cytoplasm] --> ADP [cytoplasm] + carbon dioxide [cytoplasm] + phosphoenolpyruvate [cytoplasm]	*PCK1*

p-Hydroxycinnamic acid pathway

EvolveX score	19,76 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1798	fumarate		
	r_1913	tyrosine		
	r_1993	palmitate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0032	3',5'-bisphosphate nucleotidase	adenosine 3',5'-bismonophosphate [cytoplasm] + H2O [cytoplasm] --> AMP [cytoplasm] + phosphate [cytoplasm]	*MET22*
	r_0490	glycerol-3-phosphate dehydrogenase (fad)	FAD [mitochondrion] + glycerol 3-phosphate [mitochondrion] --> dihydroxyacetone phosphate [mitochondrion] + FADH2 [mitochondrion]	*GUT2*
	r_0715	malate dehydrogenase, peroxisomal	(S)-malate [peroxisome] + NAD [peroxisome] <=> H+ [peroxisome] + NADH [peroxisome] + oxaloacetate [peroxisome]	*MDH3*
	r_4188	Glutathione-specific gamma-glutamylcyclotransferase (Gamma-GCG) (EC 4.3.2.-)	glutathione [cytoplasm] --> L-cysteinylglycine [cytoplasm] + 5-oxo-L-proline [cytoplasm]	*GCG1*

Pinocembrin pathway

EvolveX score	46,1 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1808	glycerol		
	r_1913	tyrosine		
	r_1993	palmitate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0032	3',5'-bisphosphate nucleotidase	adenosine 3',5'-bismonophosphate [cytoplasm] + H2O [cytoplasm] --> AMP [cytoplasm] + phosphate [cytoplasm]	*MET22*
	r_0363	dTMP kinase	ATP [cytoplasm] + dTMP [cytoplasm] --> ADP [cytoplasm] + dTDP [cytoplasm]	*CDC8*
	r_0735	mevalonate kinase (atp)	(R)-mevalonate [cytoplasm] + ATP [cytoplasm] --> (R)-5-phosphomevalonic acid [cytoplasm] + ADP [cytoplasm] + H+ [cytoplasm]	*ERG12*
	r_0892	phosphoglycerate kinase	1,3-bisphospho-D-glycerate [cytoplasm] + ADP [cytoplasm] <=> 3-phosphonato-D-glycerate(3-) [cytoplasm] + ATP [cytoplasm]	*PGK1*

Resveratrol pathway

EvolveX score	32,93 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1808	glycerol		
	r_1913	tyrosine		
	r_1993	palmitate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0032	3',5'-bisphosphate nucleotidase	adenosine 3',5'-bismonophosphate [cytoplasm] + H2O [cytoplasm] --> AMP [cytoplasm] + phosphate [cytoplasm]	*MET22*
	r_0735	mevalonate kinase (atp)	(R)-mevalonate [cytoplasm] + ATP [cytoplasm] --> (R)-5-phosphomevalonic acid [cytoplasm] + ADP [cytoplasm] + H+ [cytoplasm]	*ERG12*
	r_0736	mevalonate kinase (ctp)	(R)-mevalonate [cytoplasm] + CTP [cytoplasm] --> (R)-5-phosphomevalonic acid [cytoplasm] + CDP [cytoplasm] + H+ [cytoplasm]	*ERG12*
	r_0893	phosphoglycerate mutase	3-phosphonato-D-glycerate(3-) [cytoplasm] <=> 2-phospho-D-glyceric acid [cytoplasm]	*YOR283W*, *GPM1*

Vanillin pathway

EvolveX score	20,66 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1710	D-galactose		
	r_1902	methionine		
	r_1913	tyrosine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0112	acetyl-CoA synthetase	acetate [cytoplasm] + ATP [cytoplasm] + coenzyme A [cytoplasm] --> acetyl-CoA [cytoplasm] + AMP [cytoplasm] + diphosphate [cytoplasm]	*ACS1*, *ACS2*
	r_0889	phosphogluconate dehydrogenase	6-phospho-D-gluconate [cytoplasm] + NADP(+) [cytoplasm] --> carbon dioxide [cytoplasm] + D-ribulose 5-phosphate [cytoplasm] + NADPH [cytoplasm]	*GND1*, *GND2*
	r_0958	pyruvate carboxylase	ATP [cytoplasm] + bicarbonate [cytoplasm] + pyruvate [cytoplasm] --> ADP [cytoplasm] + H+ [cytoplasm] + oxaloacetate [cytoplasm] + phosphate [cytoplasm]	*PYC1*, *PYC2*
	r_1049	transketolase 1	D-xylulose 5-phosphate [cytoplasm] + ribose-5-phosphate [cytoplasm] <=> glyceraldehyde 3-phosphate [cytoplasm] + sedoheptulose 7-phosphate [cytoplasm]	*TKL1*, *TKL2*

4.4.2 EvolveXGA solutions coupling native precursor providing reactions with growth

Because EvolveXGA found a combined reaction deletion and selection niche strategy predicted to enable the use of ALE to improve the production of only 12 of the tested 28 heterologous compounds the target sets were expanded for compounds without a solution to native metabolic fluxes that provide precursors for the production pathway. The new targets were determined using SwitchX algorithm (Methods, Equation 6). With SwitchX it can be calculated which minimum set of fluxes have to change when objective is switched from growth to production. The reactions whose fluxes had to change were then used as targets in EvolveXGA. With the expanded target sets EvolveXGA could find a solution for all of the remaining heterologous pathways (Table 7). The sum of the nutrient uptakes was restricted to 75 mmol/g CDW because without the restriction the nutrient uptake became very high relative to growth for some products.

Many of the attained solutions had 2-oxoglutarate and methionine in their selection niche (i.e. solutions to growth couple target reactions identified with SwitchX for the production of 6-methylsalicylate, 8-epi-cedrol, amorpha-4-11-diene, artemisinic acid, cubebol, geraniol, hydrocortisone, patchoulol, taxa-4(20),11(12)-dien-5alpha-acetoxy-10beta-ol and valencene). In addition, the solutions had also equal or similar reaction deletions. Some of these heterologous production pathways (i.e. 8-epi-cedrol, amorpha-4-11-diene, artemisinic acid, cubebol, patchoulol, taxa-4(20),11(12)-dien-5alpha-acetoxy-10beta-ol and valencene) start with the same reactant (i.e. 2-trans,6-trans-farnesyl diphosphate) (APPENDIX 2) and have therefore very similar SwitchX identified target flux sets. Consequently, also the EvolveXGA results obtained were similar. To understand why the solutions for some other heterologous compounds were also similar, the target reaction sets were compared for 6-methylsalicylate, 8-epi-cedrol, geraniol and hydrocortisone. All these solutions had same native target reactions coupled with growth the strongest, even if the list of all targets identified using SwitchX were not identical for the different compounds.

The similar EvolveXGA solutions were thus driven with the shared targets that could be strongly coupled with growth.

Only a subset of the target reactions identified with SwitchX became growth coupled in EvolveXGA solutions. Enhancement of these reactions in adaptive laboratory evolution is expected to improve the precursor supply for production when they have a share of flux control over the production pathway. Reactions having flux control can be identified using metabolic control analysis (Wang *et al*., 2004) that requires modelling the enzyme kinetics and is beyond the scope of this work. However, the flux control of pathways is not stable but the distribution of control on the reactions changes if the enzyme properties or levels are changed. Thus, the higher amount of the precursor providing reactions become growth-coupled, the higher the likelihood of enriching production enhancing mutations in ALE.

If the growth-coupled precursor providing reactions are the same for several compounds, applying the designed ALE strategy would allow a development of a platform strain in which the precursor supply would be enhanced for all these compounds. An even better platform strain could be developed if several different production pathways have the same precursor. Liu *et al*. (2019) developed a *S. cerevisiae* platform strain with improved flux to aromatic amino acids as they are common precursors for several different compounds. Similar platform strains could be developed by setting the formation of the wanted precursor as a target reaction for EvolveXGA. Attained EvolveXGA solution could then be used during ALE to improve the flux to this precursor. After ALE this formation reaction could then be decoupled from growth and the desired synthesis pathway added.

Table 7. Solutions to growth-couple precursor providing reactions identified with SwitchX algorithm for different heterologous pathways.

6-Methylsalicylate pathway				
EvolveX score	948.71 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1586	2-oxoglutarate		
	r_1634	acetate		
	r_1902	methionine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0091	6-phosphogluconolactonase	6-O-phosphono-D-glucono-1,5-lactone [cytoplasm] + H2O [cytoplasm] --> 6-phospho-D-gluconate [cytoplasm] + H+ [cytoplasm]	*SOL3*, *SOL4*
	r_0469	glutamate decarboxylase	H+ [cytoplasm] + L-glutamate [cytoplasm] --> carbon dioxide [cytoplasm] + gamma-aminobutyrate [cytoplasm]	*GAD1*
	r_0716	malate synthase	acetyl-CoA [cytoplasm] + glyoxylate [cytoplasm] + H2O [cytoplasm] --> (S)-malate [cytoplasm] + coenzyme A [cytoplasm] + H+ [cytoplasm]	*DAL7*, *MLS1*
	r_0832	oxoglutarate dehydrogenase (lipoamide)	2-oxoglutarate [mitochondrion] + H+ [mitochondrion] + lipoamide [mitochondrion] --> carbon dioxide [mitochondrion] + S(8)-suc+cinyldihydrolipoamide [mitochondrion]	*KGD1*, *KGD2*, *LPD1*
8-Epi-cedrol pathway				
EvolveX score	1035.86 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1586	2-oxoglutarate		
	r_1683	choline		
	r_1902	methionine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0469	glutamate decarboxylase	H+ [cytoplasm] + L-glutamate [cytoplasm] --> carbon dioxide [cytoplasm] + gamma-aminobutyrate [cytoplasm]	*GAD1*
	r_0716	malate synthase	acetyl-CoA [cytoplasm] + glyoxylate [cytoplasm] + H2O [cytoplasm] --> (S)-malate [cytoplasm] + coenzyme A [cytoplasm] + H+ [cytoplasm]	*DAL7*, *MLS1*
	r_0831	oxoglutarate dehydrogenase (dihydrolipoamide S-succinyltransferase)	coenzyme A [mitochondrion] + S(8)-succinyldihydrolipoamide [mitochondrion] --> dihydrolipoamide [mitochondrion] + succinyl-CoA [mitochondrion]	*KGD1*, *KGD2*, *LPD1*
	r_0889	phosphogluconate dehydrogenase	6-phospho-D-gluconate [cytoplasm] + NADP(+) [cytoplasm] --> carbon dioxide [cytoplasm] + D-ribulose 5-phosphate [cytoplasm] + NADPH [cytoplasm]	*GND1*, *GND2*
Amorpha-4-11-diene pathway				
EvolveX score	1035.86 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1586	2-oxoglutarate		
	r_1902	methionine		
	r_2033	pyruvate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0469	glutamate decarboxylase	H+ [cytoplasm] + L-glutamate [cytoplasm] --> carbon dioxide [cytoplasm] + gamma-aminobutyrate [cytoplasm]	*GAD1*
	r_0716	malate synthase	acetyl-CoA [cytoplasm] + glyoxylate [cytoplasm] + H2O [cytoplasm] --> (S)-malate [cytoplasm] + coenzyme A [cytoplasm] + H+ [cytoplasm]	*DAL7*, *MLS1*
	r_0832	oxoglutarate dehydrogenase (lipoamide)	2-oxoglutarate [mitochondrion] + H+ [mitochondrion] + lipoamide [mitochondrion] --> carbon dioxide [mitochondrion] + S(8)-succinyldihydrolipoamide [mitochondrion]	*KGD1*, *KGD2*, *LPD1*
	r_0889	phosphogluconate dehydrogenase	6-phospho-D-gluconate [cytoplasm] + NADP(+) [cytoplasm] --> carbon dioxide [cytoplasm] + D-ribulose 5-phosphate [cytoplasm] + NADPH [cytoplasm]	*GND1*, *GND2*

Artemisinic acid pathway

EvolveX score	1034.31 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1586	2-oxoglutarate		
	r_1634	acetate		
	r_1902	methionine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0469	glutamate decarboxylase	H+ [cytoplasm] + L-glutamate [cytoplasm] --> carbon dioxide [cytoplasm] + gamma-aminobutyrate [cytoplasm]	*GAD1*
	r_0716	malate synthase	acetyl-CoA [cytoplasm] + glyoxylate [cytoplasm] + H2O [cytoplasm] --> (S)-malate [cytoplasm] + coenzyme A [cytoplasm] + H+ [cytoplasm]	*DAL7*, *MLS1*
	r_0832	oxoglutarate dehydrogenase (lipoamide)	2-oxoglutarate [mitochondrion] + H+ [mitochondrion] + lipoamide [mitochondrion] --> carbon dioxide [mitochondrion] + S(8)-succinyldihydrolipoamide [mitochondrion]	*KGD1*, *KGD2*, *LPD1*
	r_0889	phosphogluconate dehydrogenase	6-phospho-D-gluconate [cytoplasm] + NADP(+) [cytoplasm] --> carbon dioxide [cytoplasm] + D-ribulose 5-phosphate [cytoplasm] + NADPH [cytoplasm]	*GND1*, *GND2*

Beta carotene

EvolveX score	846.54 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1902	methionine		
	r_1993	palmitate		
	r_2051	spermidine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0256	catalase	2.0 hydrogen peroxide [peroxisome] --> 2.0 H2O [peroxisome] + oxygen [peroxisome]	*CTA1*
	r_0466	glucose 6-phosphate dehydrogenase	D-glucose 6-phosphate [cytoplasm] + NADP(+) [cytoplasm] --> 6-O-phosphono-D-glucono-1,5-lactone [cytoplasm] + H+ [cytoplasm] + NADPH [cytoplasm]	*ZWF1*
	r_0469	glutamate decarboxylase	H+ [cytoplasm] + L-glutamate [cytoplasm] --> carbon dioxide [cytoplasm] + gamma-aminobutyrate [cytoplasm]	*GAD1*
	r_0659	isocitrate dehydrogenase (NADP)	isocitrate [cytoplasm] + NADP(+) [cytoplasm] <=> 2-oxoglutarate [cytoplasm] + carbon dioxide [cytoplasm] + NADPH [cytoplasm]	*IDP2*

Butanol pathway

EvolveX score	401.73 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1683	choline		
	r_1761	ethanol		
	r_1902	methionine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0659	isocitrate dehydrogenase (NADP)	isocitrate [cytoplasm] + NADP(+) [cytoplasm] <=> 2-oxoglutarate [cytoplasm] + carbon dioxide [cytoplasm] + NADPH [cytoplasm]	*IDP2*
	r_0713	malate dehydrogenase	(S)-malate [mitochondrion] + NAD [mitochondrion] <=> H+ [mitochondrion] + NADH [mitochondrion] + oxaloacetate [mitochondrion]	*MDH1*
	r_0795	nucleoside diphosphate kinase	ATP [cytoplasm] + CDP [cytoplasm] --> ADP [cytoplasm] + CTP [cytoplasm]	*YNK1*
	r_1026	sulfate adenylyltransferase (ADP)	ADP [cytoplasm] + H+ [cytoplasm] + sulphate [cytoplasm] --> 5'-adenylyl sulfate [cytoplasm] + phosphate [cytoplasm]	*APA1*

Cubebol pathway				
EvolveX score	1035.86 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1586	2-oxoglutarate		
	r_1683	choline		
	r_1902	methionine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0469	glutamate decarboxylase	H+ [cytoplasm] + L-glutamate [cytoplasm] --> carbon dioxide [cytoplasm] + gamma-aminobutyrate [cytoplasm]	*GAD1*
	r_0716	malate synthase	acetyl-CoA [cytoplasm] + glyoxylate [cytoplasm] + H2O [cytoplasm] --> (S)-malate [cytoplasm] + coenzyme A [cytoplasm] + H+ [cytoplasm]	*DAL7*, *MLS1*
	r_0832	oxoglutarate dehydrogenase (lipoamide)	2-oxoglutarate [mitochondrion] + H+ [mitochondrion] + lipoamide [mitochondrion] --> carbon dioxide [mitochondrion] + S(8)-succinyldihydrolipoamide [mitochondrion]	*KGD1*, *KGD2*, *LPD1*
	r_0889	phosphogluconate dehydrogenase	6-phospho-D-gluconate [cytoplasm] + NADP(+) [cytoplasm] --> carbon dioxide [cytoplasm] + D-ribulose 5-phosphate [cytoplasm] + NADPH [cytoplasm]	*GND1*, *GND2*

Geraniol pathway				
EvolveX score	1037.33 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1586	2-oxoglutarate		
	r_1634	acetate		
	r_1902	methionine		
	reaction id	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
Deleted reactions	r_0469	glutamate decarboxylase	H+ [cytoplasm] + L-glutamate [cytoplasm] --> carbon dioxide [cytoplasm] + gamma-aminobutyrate [cytoplasm]	*GAD1*
	r_0505	glycine-cleavage complex (lipoamide)	dihydrolipoamide [mitochondrion] + NAD [mitochondrion] --> H+ [mitochondrion] + lipoamide [mitochondrion] + NADH [mitochondrion]	*GCV1*, *GCV3*
	r_0716	malate synthase	acetyl-CoA [cytoplasm] + glyoxylate [cytoplasm] + H2O [cytoplasm] --> (S)-malate [cytoplasm] + coenzyme A [cytoplasm] + H+ [cytoplasm]	*DAL7*, *MLS1*
	r_0889	phosphogluconate dehydrogenase	6-phospho-D-gluconate [cytoplasm] + NADP(+) [cytoplasm] --> carbon dioxide [cytoplasm] + D-ribulose 5-phosphate [cytoplasm] + NADPH [cytoplasm]	*GND1*, *GND2*, *LPD1*, *GCV2*, *KGD1*, *KGD2*

Hydrocortisone pathway				
EvolveX score	1064.52 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1586	2-oxoglutarate		
	r_1886	L-glucitol		
	r_1902	methionine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0466	glucose 6-phosphate dehydrogenase	D-glucose 6-phosphate [cytoplasm] + NADP(+) [cytoplasm] --> 6-O-phosphono-D-glucono-1,5-lactone [cytoplasm] + H+ [cytoplasm] + NADPH [cytoplasm]	*ZWF1*
	r_0469	glutamate decarboxylase	H+ [cytoplasm] + L-glutamate [cytoplasm] --> carbon dioxide [cytoplasm] + gamma-aminobutyrate [cytoplasm]	*GAD1*
	r_0505	glycine-cleavage complex (lipoamide)	dihydrolipoamide [mitochondrion] + NAD [mitochondrion] --> H+ [mitochondrion] + lipoamide [mitochondrion] + NADH [mitochondrion]	*GCV1*, *GCV3*
	r_0716	malate synthase	acetyl-CoA [cytoplasm] + glyoxylate [cytoplasm] + H2O [cytoplasm] --> (S)-malate [cytoplasm] + coenzyme A [cytoplasm] + H+ [cytoplasm]	*DAL7*, *MLS1*

Nicotianamine pathway

EvolveX score	459.93 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1810	glycine		
	r_1914	valine		
	r_2051	spermidine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0658	isocitrate dehydrogenase (NAD+)	isocitrate [mitochondrion] + NAD [mitochondrion] --> 2-oxoglutarate [mitochondrion] + carbon dioxide [mitochondrion] + NADH [mitochondrion]	*IDH1*, *IDH2*
	r_0959	pyruvate decarboxylase	H+ [cytoplasm] + pyruvate [cytoplasm] --> acetaldehyde [cytoplasm] + carbon dioxide [cytoplasm]	*PDC1*, *PDC5*, *PDC6*
	r_0961	pyruvate dehydrogenase	coenzyme A [mitochondrion] + NAD [mitochondrion] + pyruvate [mitochondrion] --> acetyl-CoA [mitochondrion] + carbon dioxide [mitochondrion] + NADH [mitochondrion]	*PDB1*, *PDA1*, *LPD1*, *PDX1*, *LAT1*
	r_1000	fumarate reductase (FMN)	FMNH2 [cytoplasm] + fumarate [cytoplasm] <=> FMN [cytoplasm] + succinate [cytoplasm]	*FRD1*

Patchoulol pathway

EvolveX score	1035.86 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1586	2-oxoglutarate		
	r_1683	choline		
	r_1902	methionine		
	reaction id	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
Deleted reactions	r_0469	glutamate decarboxylase	H+ [cytoplasm] + L-glutamate [cytoplasm] --> carbon dioxide [cytoplasm] + gamma-aminobutyrate [cytoplasm]	*GAD1*
	r_0716	malate synthase	acetyl-CoA [cytoplasm] + glyoxylate [cytoplasm] + H2O [cytoplasm] --> (S)-malate [cytoplasm] + coenzyme A [cytoplasm] + H+ [cytoplasm]	*DAL7*, *MLS1*
	r_0832	oxoglutarate dehydrogenase (lipoamide)	2-oxoglutarate [mitochondrion] + H+ [mitochondrion] + lipoamide [mitochondrion] --> carbon dioxide [mitochondrion] + S(8)-succinyldihydrolipoamide [mitochondrion]	*KGD1*, *KGD2*, *LPD1*
	r_0889	phosphogluconate dehydrogenase	6-phospho-D-gluconate [cytoplasm] + NADP(+) [cytoplasm] --> carbon dioxide [cytoplasm] + D-ribulose 5-phosphate [cytoplasm] + NADPH [cytoplasm]	*GND1*, *GND2*

Poly-beta-hydroxybutyrate

EvolveX score	696.21 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1683	choline		
	r_1761	ethanol		
	r_1902	methionine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0064	3-methyl-2-oxopentanoate decarboxylase	(S)-3-methyl-2-oxopentanoate [cytoplasm] + H+ [cytoplasm] --> 2-methylbutanal [cytoplasm] + carbon dioxide [cytoplasm]	*THI3*, *PDC1*, *PDC5*, *PDC6*
	r_0659	isocitrate dehydrogenase (NADP)	isocitrate [cytoplasm] + NADP(+) [cytoplasm] <=> 2-oxoglutarate [cytoplasm] + carbon dioxide [cytoplasm] + NADPH [cytoplasm]	*IDP2*
	r_0716	malate synthase	acetyl-CoA [cytoplasm] + glyoxylate [cytoplasm] + H2O [cytoplasm] --> (S)-malate [cytoplasm] + coenzyme A [cytoplasm] + H+ [cytoplasm]	*DAL7*, *MLS1*
	r_1026	sulfate adenylyltransferase (ADP)	ADP [cytoplasm] + H+ [cytoplasm] + sulphate [cytoplasm] --> 5'-adenylyl sulfate [cytoplasm] + phosphate [cytoplasm]	*APA1*

Propane-1,2-diol

EvolveX score	88.39 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1798	fumarate		
	r_1883	cysteine		
	r_1886	glucitol		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0091	6-phosphogluconolactonase	6-O-phosphono-D-glucono-1,5-lactone [cytoplasm] + H2O [cytoplasm] --> 6-phospho-D-gluconate [cytoplasm] + H+ [cytoplasm]	*SOL3*, *SOL4*
	r_0165	mitochondrial alcohol dehydrogenase	acetaldehyde [mitochondrion] + H+ [mitochondrion] + NADH [mitochondrion] --> ethanol [mitochondrion] + NAD [mitochondrion]	*ADH4*
	r_0452	fumarase, cytoplasmic	fumarate [cytoplasm] + H2O [cytoplasm] <=> (S)-malate [cytoplasm]	*FUM1*
	r_0918	phosphoserine transaminase	3-phospho-hydroxypyruvate [cytoplasm] + L-glutamate [cytoplasm] --> 2-oxoglutarate [cytoplasm] + 3-phospho-serine [cytoplasm]	*SER1*

Propane-1,3-diol

EvolveX score	438.12 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1654	ammonium		
	r_1987	ornithine		
	r_2051	spermidine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0156	alanine glyoxylate aminotransferase	glyoxylate [cytoplasm] + L-alanine [cytoplasm] --> L-glycine [cytoplasm] + pyruvate [cytoplasm]	*AGX1*
	r_0546	homoserine dehydrogenase (NADH)	H+ [cytoplasm] + L-aspartate 4-semialdehyde [cytoplasm] + NADH [cytoplasm] --> L-homoserine [cytoplasm] + NAD [cytoplasm]	*HOM6*
	r_0664	isoleucine transaminase	2-oxoglutarate [mitochondrion] + L-isoleucine [mitochondrion] <=> (S)-3-methyl-2-oxopentanoate [mitochondrion] + L-glutamate [mitochondrion]	*BAT1*
	r_0674	L-alanine transaminase	2-oxoglutarate [mitochondrion] + L-alanine [mitochondrion] <=> L-glutamate [mitochondrion] + pyruvate [mitochondrion]	*ALT1*

Taxa-4(20),11(12)-dien-5alpha-acetoxy-10beta-ol

EvolveX score	1034.31 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1586	2-oxoglutarate		
	r_1886	glucitol		
	r_1902	methionine		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0466	glucose 6-phosphate dehydrogenase	D-glucose 6-phosphate [cytoplasm] + NADP(+) [cytoplasm] --> 6-O-phosphono-D-glucono-1,5-lactone [cytoplasm] + H+ [cytoplasm] + NADPH [cytoplasm]	*ZWF1*
	r_0469	glutamate decarboxylase	H+ [cytoplasm] + L-glutamate [cytoplasm] --> carbon dioxide [cytoplasm] + gamma-aminobutyrate [cytoplasm]	*GAD1*
	r_0505	glycine-cleavage complex (lipoamide)	dihydrolipoamide [mitochondrion] + NAD [mitochondrion] --> H+ [mitochondrion] + lipoamide [mitochondrion] + NADH [mitochondrion]	*GCV1*, *GCV3*
	r_0716	malate synthase	acetyl-CoA [cytoplasm] + glyoxylate [cytoplasm] + H2O [cytoplasm] --> (S)-malate [cytoplasm] + coenzyme A [cytoplasm] + H+ [cytoplasm]	*DAL7*, *MLS1*

Valencene

EvolveX score	1035.86 mmol/g CDW			
Selection niche	**reaction id**	**compound name**		
	r_1586	2-oxoglutarate		
	r_1902	methionine		
	r_2033	pyruvate		
Deleted reactions	**reaction id**	**reaction name**	**reaction**	**gene(s) encoding for the enzyme(s) catalyzing the reaction**
	r_0469	glutamate decarboxylase	H+ [cytoplasm] + L-glutamate [cytoplasm] --> carbon dioxide [cytoplasm] + gamma-aminobutyrate [cytoplasm]	*GAD1*
	r_0716	malate synthase	acetyl-CoA [cytoplasm] + glyoxylate [cytoplasm] + H2O [cytoplasm] --> (S)-malate [cytoplasm] + coenzyme A [cytoplasm] + H+ [cytoplasm]	*DAL7*, *MLS1*
	r_0831	oxoglutarate dehydrogenase (dihydrolipoamide S-succinyltransferase)	coenzyme A [mitochondrion] + S(8)-succinyldihydrolipoamide [mitochondrion] --> dihydrolipoamide [mitochondrion] + succinyl-CoA [mitochondrion]	*KGD1*, *KGD2*, *LPD1*
	r_0889	phosphogluconate dehydrogenase	6-phospho-D-gluconate [cytoplasm] + NADP(+) [cytoplasm] --> carbon dioxide [cytoplasm] + D-ribulose 5-phosphate [cytoplasm] + NADPH [cytoplasm]	*GND1*, *GND2*

5 Conclusions

EvolveXGA was able to find growth-coupling strategies including reaction deletions and selection niche predicted to allow using ALE for improving production of several different compounds by *S. cerevisiae*. Therefore, the results supported the primary hypothesis set for the demonstration of EvolveXGA algorithm. One of the strategies found by EvolveXGA for growth-coupling the reactions of the oxalate pathway for glycolic acid synthesis in *S. cerevisiae* included an intuitive solution derived by a domain expert scientist. In addition, several other strategies leading into stronger growth-coupling of the oxalate pathway reactions than the intuitive solution were found. This shows that even if intuitive design would be possible, modelling can allow finding better solutions.

The secondary hypothesis was that EvolveXGA would be generalizable across different products and find strategies to couple the corresponding heterologous production pathways with growth. This hypothesis was not fully met with all analysed 29 heterologous products to *S. cerevisiae* since with the tested set of four reaction deletions EvolveXGA could find growth-coupling strategies for only 13 production pathways including the oxalate pathway for glycolic acid synthesis. However, it is possible that EvolveXGA could have found growth-coupling strategies also for the remaining pathways if more reaction deletions were allowed in the search.

Increasing the number of reaction deletions would, nevertheless, increase the possibility that realizing one of the deletions would not be possible *in vivo*. Even if EvolveXGA could not suggest deletions to growth essential reactions the essentiality was not automatically incorporated on gene level. Some solutions were discarded in the evaluation as realizing the reaction deletions in the solution would have required deleting a gene that encoded an enzyme that is also needed for growth. To avoid this challenge that arises from enzyme promiscuity (i.e. enzymes that are associated with several reactions) the search for growth-coupling strategies could be performed directly on gene deletions as this would remove solutions where genes encoding growth essential enzymes would be deleted. Machado *et al*. (2016) showed that this could be done by using a stoichiometric matrix in the simulations that includes gene-protein-reaction associations and no additional changes would be needed in the EvolveXGA algorithm.

Since growth-coupling strategies were not found for some reactions of each heterologous production pathway, the target reaction sets for EvolveXGA were expanded to native reactions providing precursors for production. Such reactions were identified as those whose flux increase would be needed to enhance the production. When these reactions were used as EvolveXGA targets a growth-coupling solution was found for all the remaining products. The found strategies are predicted to enable using ALE for directing cellular resources better towards the desired production.

Thus, the results show that it is possible to find growth-coupling strategies for a wide variety of target reactions *in silico* with this novel approach combining selection niche with reaction deletions. Furthermore, the extension of searching a selection niche in addition to reaction deletions had a notable impact. Solutions with stronger growth-coupling for the three first reactions of oxalate pathway for glycolic acid synthesis (i.e. oxaloacetase, oxalate-CoA ligase and glyoxylate dehydrogenase) were attained when reaction deletions were combined with selection niche than when only reaction deletions were searched. In addition, for some pathways (e.g. naringenin pathway) growth-coupling was mainly created with selection niche nutrients. Evaluation of the growth-coupling strategies also showed that some of these strategies have potential to be validated *in vivo*. However, further literature based research about the effect of the required gene deletions on cell viability is needed to rank the strategies to be implemented in the wet lab. It is possible that the best solutions to be implemented *in vivo* are not among the final local optimal solutions but can be found among the sub-optimal solutions. In addition, also other variable values, especially higher number of reaction deletions, could yield more potential solutions. Furthermore, in addition to the nutritional environment and metabolic network modifications also the level of available oxygen affects how the fluxes are distributed in the metabolic network. In this work aerobic conditions were considered, but it would also be possible to search growth-coupling strategies for oxygen limited and anaerobic conditions using EvolveXGA by constraining the oxygen exchange in the metabolic model.

EvolveXGA provides means to accelerate the production strain improvement towards industrially feasible production by enabling the use of ALE. ALE is simple to perform and tackles the strain improvement challenge of unknown complex regulation behind phenotypes. EvolveXGA was found generalizable across various industrially relevant compounds and has thus potential to facilitate development of production strains. As synthesizing chemicals with microbial fermentation allows the use of sustainable resources, EvolveXGA can make an important contribution in the battle against climate warming and resource sufficiency.

References

Asadollahi M. A., Maury J., Patil K. R., Schalk M., Clark A. and Nielsen J., (2009), Enhancing sesquiterpene production in *Saccharomyces cerevisiae* through *in silico* driven metabolic engineering, *Metab. Eng.* **11** 328-334.

Becker-Kettern J., Paczia N., Conrotte J.-F., Kay D. P., Guignard C., Jung P. P. and Linster C., (2016), *Saccharomyces cerevisiae* forms D-2-hydroxyglutarate and couples its degradation to D-lactate formation via a cytosolic transhydrogenase, *The journal of biological chemistry* **291**(12) 6036-6058.

Brochado A. R., Matos C., Møller B. L., Hansen J., Mortensen U. H. and Patil K. R., (2010), Improved vanillin production in baker's yeast through *in silico* design, *Microb. Cell Fact.* **9**:84.

Bugard A. P., Vaidyaraman S. and Maranas C. D., (2001), Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments, *Biotechnol. Prog.* **17** 791-797.

Bugard, A.P., Pharkya P. and Maranas D., (2003), OptKnock: A-bilevel programming framework for identifying gene knockout strategies for microbial strain optimization, *Biotechnol. Bioeng.* **84(6)** 647-657.

Campodonico M. A., Andrews B. A., Asenjo J. A., Palsson B. O. and Feist A. M., (2014), Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path, *Metab. Eng.* **25** 140-158.

Castañeda M. T., Nuñez S., Voget C. and De Battista H., (2019), *In silico* optimization of lipid production in *Rhodosporidium toruloides* by gene knockout strategies, *IFAC PapersOnLine* **52**(1) 94-99.

Castillo S., Patil K.R. and Jouhten P., (2019), Yeast Genome-Scale Metabolic Models for Simulating Genotype–Phenotype Relations. In: Sá-Correia I. (eds) *Yeasts in Biotechnology and Human Health*. *Progress in Molecular and Subcellular Biology **58***, Springer, Cham.

Chen Y., Siewers V. and Nielsen J., (2012), Profiling of cytosolic and peroxisomal acetyl-CoA metabolism in *Saccharomyces cerevisiae*, *PLoS One* **7**(8) e42475.

Cotten C. and Reed J. L., (2013), Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering, *Biotechnol. J.* **8** 595-604.

DeLuna A., Avendaño A., Riego L. and González A., (2001), NADP-glutamate dehydrogenase isoenzymes of *Saccharomyces cerevisiae*, *The journal of biological chemistry* **276**(47) 43775-43783.

Dragosits M. and Mattanovich D., (2013), Adaptive laboratory evolution - principles and applications for biotechnology, *Microb. Cell Fact.* **12**(64).

Edwards J. S., Ibarra R. U. and Palsson B. O., (2001), *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data, *Nat. Biotechnol.* **19** 125-130.

Edwards J. S., Covert M. and Palsson B., (2002), Metabolic modelling of microbes: the flux-balance approach, *Environ. Microbiol.* **4**(3) 133-140.

Elgersma Y., van Roermund W.T., Wanders R. J. A. and Tabak H. F., (1995), Peroxisomal and mitochondrial carnitine acetyltransferases of *Saccharomyces cerevisiae* are encoded by a single gene, *The EMBO Journal* **14** 3472-3479.

Erb T. J., Jones P. R. and Bar-Even A., (2017), Synthetic metabolism: metabolic engineering meets enzyme design, *Curr. Opin. Chem. Biol.* **37** 56-62.

Fakas S., (2017), Lipid biosynthesis in yeasts: A comparison of the lipid biosynthetic pathway between the model nonoleaginous yeast *Saccharomyces cerevisiae* and the model oleaginous yeast *Yarrowia lipolytica*, *Eng. Life Sci.* **17** 292-302.

Feist A. M., Zielinski D. C., Orth J. D., Schellenberger J., Herrgard M. J. and Palsson B. Ø., (2010), Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*, *Metab. Eng.* **12** 173-186.

Fernández E., Moreno F. and Rodicio R., (1992), The ICL1 gene from *Saccharomyces cerevisiae*, *Eur. J. Biochem.* **204** 983-990.

Flikweert M. T., de Swaaf M., van Dijken J. P. and Pronk J. T., (1999), Growth requirements of pyruvate-decarboxylase-negative *Saccharomyces cerevisiae*, *FEMS microbiology letters* **174** 73-79.

Fong S. S., Bugard A. P., Herring C. D., Knight E. M., Blattner F. R., Maranas C. D. and Palsson B. O., (2005), In silico design and adaptive evolution of Escherichia coli for production of lactic acid, *Biotechnol. Bioeng.* **91**(5) 643-648.

Graybill E. R., Rouhier M. F., Kirby C. E. and Hawes J. W., (2007), Functional comparison of citrate synthase isoforms from S*. cerevisiae*, *Arch. Biochem. Biophys.* **465** 26-37.

Gu C., Kim G. B., Kim W. J., Kim H. U. and Lee S. Y., (2019), Current status and applications of genome scale metabolic models, *Genome Biol.* **20** 121.

Haselbeck R. J. and McAlister-Henn, (1993), Function and expression of yeast mitochondrial NAD- and NADP-specific isocitrate dehydrogenases, *The journal of biological chemistry* **266**(16) 12116-12122.

Hazelwood L. A., Daran J-M., van Maris A. J. A, Pronk J. T. and Dickinson J. R., (2008), The Erlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism, *Appl. Environ. Microbiol.* **74**(8) 2259-2266.

Hohmann S. and Cederberg H., (1990), Autoregulation may control the expression of yeast pyruvate decarboxylase structural genes PDC1 and PDC5, *Eur. J. Biochem.* **188** 615-621.

Hohmann S., (1991), Characterization of PDC6, a third structural gene for pyruvate decarboxylase in *Saccharomyces cerevisiae*, *J. Bacteriol.* **173** 7963-7969.

Ibarra R. U., Edwards J. S. and Palsson B. O., (2002), E*scherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth, *Nature* **420** 186-189.

Iranmanesh E., Asadollahi M. A. and Biria D., (2020), Improving L-phenylacetylcarbinol production in *Saccharomyces cerevisiae* by *in silico* aided metabolic engineering, *J. Biotechnol.* **308** 27-34.

Jouhten P., Boruta T., Andrejev S., Pereira F., Rocha I. and Patil K. R., (2016), Yeast metabolic chassis designs for diverse biotechnological products, *Sci. Rep.* **6**:29694.

Jouhten P., Konstantinidis D., Pereira F., Andrejev S., Grkovska K., Ghiachi P., Beltran G., Almaas E., Mas A., Warringer J., Gonzalez R., Morales P. and Patil K. R., (2021 Preprint), Predictive evolution of metabolic phenotypes using model-designed selection niches, bioRxiv 2021.05.14.443989;
doi: https://doi.org/10.1101/2021.05.14.443989.

Jouhten P. T. and Patil K. R., (2016), Methods for adaptive laboratory evolution, WO/2018/007522.

Kauffman K. J., Prakash P. and Edwards J. S., (2003), Advances in flux balance analysis, *Curr. Opin. Biotechnol.* **14** 491-496.

Keys D. A. and McAlister-Henn L., (1990), Subunit structure, expression, and function of NAD(H)-specific isocitrate dehydrogenase in *Saccharomyces cerevisiae, J. Bacteriol.* **172**(8) 4280-4287.

Kim J., Reed J. L. and Maravelias C. T., (2011), Large-scale bi-level strain design approaches and mixed-integer programming solution techniques, *PLoS One* **6**(9) e24162.

Kispal G., Rosenkrantz M., Guarente L. and Srere P. A., (1988), Metabolic changes in S*accharomyces cerevisiae* strain lacking citrate synthases, *The journal of biological chemistry* **263**(23) 11145-11149.

Krivoruchko A., Zhang Y., Siewers V., Chen Y. and Nielsen J., (2015), Microbial acetyl-CoA metabolism and metabolic engineering, *Metab. Eng.* **28** 28-42.

Lawson J. E., Behal R. H. and Reed L. J., (1991a), Disruption and mutagenesis of the *Saccharomyces cerevisiae* PDX1 gene encoding the protein X component of the pyruvate dehydrogenase complex, *Biochem.* **30** 2834-2839.

Lawson J. E., Niu X. D. and Reed L. J., (1991b), Functional analysisof the domains of dihydrolipoamide acetyltransferase from *Saccharomyces cerevisiae*, *Biochem.* **30** 11249-11254.

Lee S. Y., Kim H. U., Chae T. U., Cho J. S., Kim J. W., Shin J. H., Kim D. I., Ko Y-S., Jang W. D. and Jang Y-S., (2019), A comprehensive metabolic map for production of bio-based chemicals, *Nat. Catal.* **2** 18-33.

Lewis N. E., Hixson K. K., Conrad T. M., Lerman J. A., Charusanti P., Polpitiya A. D., Adkins J. N., Schramm G., Purvine S. O., Lopez-Ferrer D., Weitz K. K., Eils R., König R., Smith R. D. and Palsson B. Ø., (2010), Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models, *Mol. Syst. Biol.* **6**:390.

Liu J-Q., Nagata S., Dairi T., Misoni H., Shimizu S. and Yamada H., (1997), The GLY1 gene of *Saccharomyces cerevisiae* encodes a low-specific L-threonine aldolase that catalyzes cleavage of L-*allo*-threonine and L-threonine to glycine – Expression of gene in *Escherichia coli* and purification and characterization of the enzyme, *Eur. J. Biochem.* **245** 289-293.

Liu Q., Yu T., Li X., Chen Y., Campbell K., Nielsen J. and Chen Y., (2019), Rewiring carbon metabolism in yeast for high level production of aromatic chemicals, *Nat. Commun.* **10**(1) 4976.

Lu H., Li F., Sánchez B. J., Zhu Z., Li G., Domenzain I., Marcišauskas S., Anton P. M., Lappa D., Lieven C., Beber M. E., Sonnenschein N., Kerkhoven J. and Nielsen J., (2019), A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism, *Nat. Commun.* **10** 3586.

Luttik M. A. H., Kötter P., Salomons F. A., van der Klei I. J., van Dijken J. P. and Pronk J. T., (2000), The *Saccharomyces cerevisiae* ICL2 gene encodes a mitochondrial 2-methylisocitrate lyase involved in propionyl-coenzyme A metabolism, *J. Bacteriol.* **182**(24), 7007-7013.

Machado D., Herrgård M. J. and Rocha I., (2016), Stoichiometric representation of gene-protein-reaction associations leverages constraint based analysis from reaction to gene-level phenotype prediction, *PLoS Comput. Biol.* **12**(10) e1005140.

Mahadevan R. and Schilling C. H., (2003), The effects of alternate optimal solutions in constraint-based genome-scale metabolic models, *Metab. Eng.* **5** 264-276.

Maia P., Rocha M. and Rocha I., (2016), *In silico* constraint-based strain optimization methods: the quest for optimal cell factories, *Microbiol. Mol. Biol. Rev.* **80**(1) 45-67.

Manzer L. E., van der Waal J. C. and Imhof P., (2013), The industrial playing field for the conversion of biomass to renewable fuels and chemicals, In: *Catalytic process development for renewable materials,* ed. Imhof P. and van der Waal J. C., 1st edition, Wiley-VCH Verlag GmbH & Co.

McNeil J. B., McIntosh E. M., Taylor B. V., Zhang F-r., Tang S. and Bognar A. L., (1994), Cloning and molecular characterization of three genes including two genes encoding serine hydroxymethyltransferases, whose inactivation is required to render yeast auxotrophic for glycine, *The journal of biological chemistry* **269**(12) 9155-9165.

Miran S. G., Lawson J. E. and Reed L. J., (1993), Characterization of PDHβ1, the structural gene for the pyruvate dehydrogenase β subunit from *Saccharomyces cerevisiae*, *Proc. Natl. Acad. Sci. U.S.A.* **90** 1252-1256.

Moscatello N. and Pfeifer B. A., (2018), Constraint-based metabolic targets for the improved production of heterologous compounds across molecular classification, *AIChE Journal* **64**(12) 4208-4217.

Navarro-Aviño J. P., Miralles R. P. V. J. and Serrano R., (1999), A proposal for nomenclature of aldehyde dehydrogenases in *Saccharomyces cerevisiae* and characterization of the stress-inducible ALD2 and ALD3 genes, *Yeast* **15**(10A) 829-842.

Najafpour G. D., (2015), *Biochemical engineering and biotechnology,* 2. edition, Elsevier, Amsterdam, 652 p.

Niyonzima F. N., (2019), Production of microbial industrial enzymes, *Acta scientific microbiology* **2**(12) 75-89.

O'Brien E. J., Monk J. M. and Palsson B. O., (2015), Using genome-scale models to predict biological capabilities, *Cell* **161** 971-987.

Ohta E., Nakayama Y., Mukai Y., Bamba T. and Fukusaki E., (2016), Metabolomic approach for improving ethanol stress tolerance in S*accharomyces cerevisiae*, *J. Biosci. Bioeng.* **121**(4) 399-405.

Orth J. D., Thiele I. and Palsson B. Ø., (2010), What is flux balance analysis?, *Nat. Biotechnol.* **28**(3) 245-248.

Otero J. M., Cimini D., Patil K. R., Poulsen S. G., Olsson L. and Nielsen J., (2013), Industrial systems biology of *Saccharomyces cerevisiae* enables novel succinic acid cell factory, *PLoS One* **8**(1) e54144.

Pál C. and Papp B., (2017), Evolution of complex adaptions in molecular systems, *Nat. Ecol. Evol.* **1**(8) 1084-1092.

Park J. M., Park H. M., Kim W. J., Kim H. U., Kim T. Y. and Lee S. Y., (2012), Flux variability scanning based on enforced objective flux for identifying gene amplification targets, *BMC Syst. Biol.* **6**(106).

Patil K. R., Rocha I., Förster J. and Nielsen J., (2005), Evolutionary programming as a platform for *in silico* metabolic engineering, *BMC Bioinf.* **6**:308.

Petrik M., Käppeli O. and Fiechter A., (1983), An expanded concept for the glucose effect in the yeast *Saccharomyces uvarum*: involvement of short- and long-term regulation, *J. Gen. Microbiol.* **129** 43-49.

Pharkya P., Bugard A. P. and Maranas C. D., (2003), Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock, *Biotechnol. Bioeng.* **84**(7) 887-899.

Pharkya P., Burgard A. P. and Maranas C. D., (2004), OptStrain: a computational framework for redesign of microbial production systems, *Genome Res.* **14** (11) 2367-2376.

Pharkya P. and Maranas C. D., (2006), An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems, *Metab. Eng.* **8** 1-13.

Pronk J. T., Wenzel T. J., Luttik M. A. H., Klaassen C. C. M., Scheffers A., Steensma H. Y. and van Dijken J. P., (1994), Energetic aspects of glucose metabolism in a pyruvate-dehydrogenase-negative mutant of *Saccharomyces cerevisiae*, *Microbiology* **140** 601-610.

Pronk J. T., Steensma H. Y. and van Dijken J. P., (1996), Pyruvate metabolism in *Saccharomyces cerevisiae, Yeast* **12** 1607-1633.

Ranganathan S., Suthers P. F. and Maranas C. D., (2010), OptForce: An Optimization Procedure for identifying all genetic manipulations leading to targeted overproductions, *PLoS Comput. Biol.* **6**(4) e1000744.

Ranganathan S., Tee T. W., Chowdhury A., Zomorrodi A. R., Yoon J. M., Fu Y., Shanks J. V. and Maranas C. D., (2012), An integrated computational and experimental study for overproducing fatty acids in *Escherichia coli*, *Metab. Eng.* **14** 687-704.

Reyes L. H., Gomez J. M. and Kao K. C., (2014), Improving carotenoids production in yeast via adaptive laboratory evolution, *Metab. Eng.* **21** 26-33.

Sarker R. A. and Kazi M. F. A., (2003), Population size, search space and quality of solution: an experimental study, *The 2003 Congress on Evolutionary Computation*, CEC '03., Canberra, **3** 2011-2018.

Sánchez B., Li F., Lu H. and Kerkhoven E., (2020), SysBioChalmers/yeast-GEM: yeast 8.3.5 (Version v8.3.5), Zenodo, http://doi.org/10.5281/zenodo.3738008.

Sandberg T. E., Salazar M. J., Weng L. L., Palsson B. O. and Feist A. M., (2019), The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology, *Metab. Eng.* **56** 1-16.

Segrè D., Vitkup D. and Church G. M., (2002), Analysis of optimality in natural and perturbed metabolic networks, *PNAS* **99**(23) 15112-15117.

Shlomi T., Berkman O. and Ruppin E., (2005), Regulatory on/off minimization of metabolic flux changes after genetic pertubations, *PNAS* **102** (21) 7695-7700.

Smith J. M. and Haigh J., (1974), The hitch-hiking effect of favourable gene, *Genet. Res. Camb.* **23** 23-35.

Suthers P. F., Dinh H. V., Fatma Z., Shen Y., Chan S. H. J., Rabinowitz J. D., Zhao H. and Maranas C. D., (2020), Genome-scale metabolic reconstruction of the non-model yeast *Issatchenkia orientalis* SD108 and its application to organic acids production, *Metabolic engineering communications* **11** e00148.

Szappanos B., Fritzemeier J., Csörgő B., Lázár V., Lu X., Fekete G., Bálint B., Herczeg R., Nagy I., Notebaart, R. A., Lercher M. J., Pál C. and Papp B., (2016), Adaptive evolution of complex innovations through stepwise metabolic niche expansion, *Nat. Commun.* **7:**11607.

Tepper N. and Shlomi T., (2010), Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways, *Bioinformatics* **26**(4) 536-543.

Toivari M., Ilmén M. and Penttilä M., (2017), Improved production of oxalyl-coA, glyoxylate and/or glycolic acid, WO 2019/020870.

Thomas D. and Surdin-Kerjan Y., (1997), Metabolism of sulfur amino acids in *Saccharomyces cerevisiae*, *Microbiol. Mol. Biol. Rev.* **61**(4) 503-532.

van Urk H., Mak P. R., Scheffers A. and van Dijken J. P., (1988), Metabolic responses of *Saccharomyces cerevisiae* CBS 8066 and *Candida utilis* CBS 621 upon transition from glucose limitation to glucose excess, *Yeast* **4** 283-291.

Varma A., Boesch B. W. and Palsson B. O., (1993), Biochemical production capabilities of *Escherichia coli*, *Biotechnol. Bioeng.* **42** 59-73

Varma A. and Palsson B. O., (1994), Metabolic flux balancing: Basic concepts, scientific and practical use, *Nat. Biotechnol.* **12** 994-998.

Vohra M., Manwar J., Manmode R., Padgilwar S. and Patil S., (2014), Bioethanol production: Feedstock and current technologies, *Engineering* **2**(1) 573-584.

Wang L., Birol I., Hatzimanikatis V., (2004), Metabolic control analysis under uncertainty: Framework development and case studies, *Biophys. J.* **87** 3750-3763.

Wehrs M., Tanjore D., Eng T., Lievense J., Pray T. R. and Mukhopadhyay, (2019), Engineering robust production microbes for large-scale cultivation, *Trends Microbiol.* **27**(6) 524-537.

Wenzel T. J., van den Berg M. A., Visser W., van den Berg J. A. and Steensma H. Y., (1992), Characterization of *Saccharomyces cerevisiae* mutants lacking the E1α subunit of the pyruvate dehydrogenase complex, *Eur. J. Biochem.* **209** 697-705.

Xu P., Ranganathan S., Fowler Z. L., Maranas C. D. and Koffas M. A. G., (2011), Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA, *Metab. Eng.* **13** 578-587.

Yang L., Cluett W. R. and Mahadevan R., (2011), EMILiO: A fast algorithm for genome-scale strain design, *Metab. Eng.* **13** 272-281.

Zabranska J. and Pokorna D., (2018), Bioconversion of carbon dioxide to methane using hydrogen and hydrogenotrophic methanogens, *Biotechnol. Adv.* **36(3)** 707-720.

Zhang X., Tervo C. J. and Reed, J. L., (2016), Metabolic assessment of *E. coli* as a biofactory for commercial products, *Metab. Eng.* **35** 64-74.

APPENDIX 1. Modifications to reaction lower and upper bounds in the used *S. cerevisiae* genome-scale metabolic model. The following lower and upper bounds were set to 0.

reaction id	reaction name	reaction	bound
r_0399	fatty-acid--CoA ligase (decanoate)	ATP [peroxisome] + coenzyme A [peroxisome] + decanoate [peroxisome] <=> AMP [peroxisome] + decanoyl-CoA [peroxisome] + diphosphate [peroxisome]	lb
r_0400	fatty-acid--CoA ligase (dodecanoate)	ATP [peroxisome] + coenzyme A [peroxisome] + laurate [peroxisome] <=> AMP [peroxisome] + diphosphate [peroxisome] + lauroyl-CoA [peroxisome]	lb
r_0402	fatty-acid--CoA ligase (hexadecanoate)	ATP [peroxisome] + coenzyme A [peroxisome] + palmitate [peroxisome] <=> AMP [peroxisome] + diphosphate [peroxisome] + palmitoyl-CoA [peroxisome]	lb
r_0410	fatty-acid--CoA ligase (octanoate)	ATP [peroxisome] + coenzyme A [peroxisome] + octanoate [peroxisome] <=> AMP [peroxisome] + diphosphate [peroxisome] + octanoyl-CoA [peroxisome]	lb
r_0412	fatty-acid--CoA ligase (tetradecanoate)	ATP [peroxisome] + coenzyme A [peroxisome] + myristate [peroxisome] <=> AMP [peroxisome] + diphosphate [peroxisome] + myristoyl-CoA [peroxisome]	lb
r_0446	formate-tetrahydrofolate ligase	ATP [cytoplasm] + formate [cytoplasm] + THF [cytoplasm] <=> 10-formyl-THF [cytoplasm] + ADP [cytoplasm] + phosphate [cytoplasm]	lb
r_0487	glycerol dehydrogenase (NADP-dependent)	glycerol [cytoplasm] + NADP(+) [cytoplasm] --> glycerone [cytoplasm] + H+ [cytoplasm] + NADPH [cytoplasm]	ub
r_2194	fatty-acid--CoA ligase (dodecanoate), ER membrane	coenzyme A [endoplasmic reticulum membrane] + ATP [endoplasmic reticulum membrane] + laurate [endoplasmic reticulum membrane] <=> lauroyl-CoA [endoplasmic reticulum membrane] + AMP [endoplasmic reticulum membrane] + diphosphate [endoplasmic reticulum membrane]	lb
r_2195	fatty-acid--CoA ligase (tetradecanoate), ER membrane	coenzyme A [endoplasmic reticulum membrane] + ATP [endoplasmic reticulum membrane] + myristate [endoplasmic reticulum membrane] <=> myristoyl-CoA [endoplasmic reticulum membrane] + AMP [endoplasmic reticulum membrane] + diphosphate [endoplasmic reticulum membrane]	lb
r_2196	fatty-acid--CoA ligase (hexadecanoate), ER membrane	coenzyme A [endoplasmic reticulum membrane] + ATP [endoplasmic reticulum membrane] + palmitate [endoplasmic reticulum membrane] <=> palmitoyl-CoA [endoplasmic reticulum membrane] + AMP [endoplasmic reticulum membrane] + diphosphate [endoplasmic reticulum membrane]	lb
r_2197	fatty-acid--CoA ligase (hexadecenoate), ER membrane	coenzyme A [endoplasmic reticulum membrane] + ATP [endoplasmic reticulum membrane] + palmitoleate [endoplasmic reticulum membrane] <=> palmitoleoyl-CoA(4-) [endoplasmic reticulum membrane] + AMP [endoplasmic reticulum membrane] + diphosphate [endoplasmic reticulum membrane]	lb
r_2198	fatty-acid--CoA ligase (octadecanoate), ER membrane	coenzyme A [endoplasmic reticulum membrane] + ATP [endoplasmic reticulum membrane] + stearate [endoplasmic reticulum membrane] <=> stearoyl-CoA [endoplasmic reticulum membrane] + AMP [endoplasmic reticulum membrane] + diphosphate [endoplasmic reticulum membrane]	lb
r_2199	fatty-acid--CoA ligase (octadecenoate), ER membrane	coenzyme A [endoplasmic reticulum membrane] + ATP [endoplasmic reticulum membrane] + oleate [endoplasmic reticulum membrane] <=> oleoyl-CoA [endoplasmic reticulum membrane] + AMP [endoplasmic reticulum membrane] + diphosphate [endoplasmic reticulum membrane]	lb
r_2200	fatty-acid--CoA ligase (dodecanoate), lipid particle	coenzyme A [lipid particle] + ATP [lipid particle] + laurate [lipid particle] <=> diphosphate [lipid particle] + AMP [lipid particle] + lauroyl-CoA [lipid particle]	lb
r_2201	fatty-acid--CoA ligase (tetradecanoate), lipid particle	coenzyme A [lipid particle] + ATP [lipid particle] + myristate [lipid particle] <=> diphosphate [lipid particle] + AMP [lipid particle] + myristoyl-CoA [lipid particle]	lb
r_2202	fatty-acid--CoA ligase (hexadecanoate), lipid particle	coenzyme A [lipid particle] + ATP [lipid particle] + palmitate [lipid particle] <=> diphosphate [lipid particle] + AMP [lipid particle] + palmitoyl-CoA [lipid particle]	lb
r_2203	fatty-acid--CoA ligase (hexadecenoate), lipid particle	coenzyme A [lipid particle] + ATP [lipid particle] + palmitoleate [lipid particle] <=> diphosphate [lipid particle] + AMP [lipid particle] + palmitoleoyl-CoA(4-) [lipid particle]	lb
r_2204	fatty-acid--CoA ligase (octadecanoate), lipid particle	coenzyme A [lipid particle] + ATP [lipid particle] + stearate [lipid particle] <=> diphosphate [lipid particle] + AMP [lipid particle] + stearoyl-CoA [lipid particle]	lb
r_2205	fatty-acid--CoA ligase (octadecenoate), lipid particle	coenzyme A [lipid particle] + ATP [lipid particle] + oleate [lipid particle] <=> diphosphate [lipid particle] + AMP [lipid particle] + oleoyl-CoA [lipid particle]	lb
r_2206	fatty-acid--CoA ligase (hexadecenoate), peroxisome	ATP [peroxisome] + coenzyme A [peroxisome] + palmitoleate [peroxisome] <=> AMP [peroxisome] + diphosphate [peroxisome] + palmitoleoyl-CoA(4-) [peroxisome]	lb
r_2207	fatty-acid--CoA ligase (octadecanoate), peroxisome	ATP [peroxisome] + coenzyme A [peroxisome] + stearate [peroxisome] <=> AMP [peroxisome] + diphosphate [peroxisome] + stearoyl-CoA [peroxisome]	lb
r_2208	fatty-acid--CoA ligase (octadecenoate), peroxisome	ATP [peroxisome] + coenzyme A [peroxisome] + oleate [peroxisome] <=> AMP [peroxisome] + diphosphate [peroxisome] + oleoyl-CoA [peroxisome]	lb
r_2209	fatty-acid--CoA ligase (arachidate), cell envelope	ATP [cell envelope] + coenzyme A [cell envelope] + arachidate [cell envelope] <=> AMP [cell envelope] + diphosphate [cell envelope] + icosanoyl-CoA [cell envelope]	lb
r_2210	fatty-acid--CoA ligase (behenate), cell envelope	ATP [cell envelope] + coenzyme A [cell envelope] + behenate [cell envelope] <=> AMP [cell envelope] + diphosphate [cell envelope] + docosanoyl-CoA [cell envelope]	lb
r_2211	fatty-acid--CoA ligase (lignoceric acid), cell envelope	ATP [cell envelope] + coenzyme A [cell envelope] + lignoceric acid [cell envelope] <=> AMP [cell envelope] + diphosphate [cell envelope] + tetracosanoyl-CoA [cell envelope]	lb
r_2212	fatty-acid--CoA ligase (cerotic acid), cell envelope	ATP [cell envelope] + coenzyme A [cell envelope] + cerotic acid [cell envelope] <=> AMP [cell envelope] + diphosphate [cell envelope] + hexacosanoyl-CoA [cell envelope]	lb
r_2213	fatty-acid--CoA ligase (behenate), ER membrane	coenzyme A [endoplasmic reticulum membrane] + ATP [endoplasmic reticulum membrane] + behenate [endoplasmic reticulum membrane] <=> docosanoyl-CoA [endoplasmic reticulum membrane] + AMP [endoplasmic reticulum membrane] + diphosphate [endoplasmic reticulum membrane]	lb
r_2214	fatty-acid--CoA ligase (lignoceric acid), ER membrane	coenzyme A [endoplasmic reticulum membrane] + ATP [endoplasmic reticulum membrane] + lignoceric acid [endoplasmic reticulum membrane] <=> tetracosanoyl-CoA [endoplasmic reticulum membrane] + AMP [endoplasmic reticulum membrane] + diphosphate [endoplasmic reticulum membrane]	lb
r_2215	fatty-acid--CoA ligase (cerotic acid), ER membrane	coenzyme A [endoplasmic reticulum membrane] + ATP [endoplasmic reticulum membrane] + cerotic acid [endoplasmic reticulum membrane] <=> hexacosanoyl-CoA [endoplasmic reticulum membrane] + AMP [endoplasmic reticulum membrane] + diphosphate [endoplasmic reticulum membrane]	lb
r_2216	fatty-acid--CoA ligase (behenate), lipid particle	coenzyme A [lipid particle] + ATP [lipid particle] + behenate [lipid particle] <=> diphosphate [lipid particle] + AMP [lipid particle] + docosanoyl-CoA [lipid particle]	lb
r_2217	fatty-acid--CoA ligase (lignoceric acid), lipid particle	coenzyme A [lipid particle] + ATP [lipid particle] + lignoceric acid [lipid particle] <=> diphosphate [lipid particle] + AMP [lipid particle] + tetracosanoyl-CoA [lipid particle]	lb
r_2218	fatty-acid--CoA ligase (cerotic acid), lipid particle	coenzyme A [lipid particle] + ATP [lipid particle] + cerotic acid [lipid particle] <=> diphosphate [lipid particle] + AMP [lipid particle] + hexacosanoyl-CoA [lipid particle]	lb
r_4158	NADPH2:quinone oxidoreductase	H+ [cytoplasm] + NADPH [cytoplasm] + 2.0 quinone [cytoplasm] <=> NADP(+) [cytoplasm] + 2.0 1,4-benzosemiquinone [cytoplasm]	lb
r_4198	Benzil reductase ((S)-benzoin forming) IRC24 (EC 1.1.1.320) (Increased recombination centers protein 24)	NADP(+) [cytoplasm] + (S)-benzoin [cytoplasm] <=> H+ [cytoplasm] + NADPH [cytoplasm] + benzil [cytoplasm]	ub
r_4275	Fe(II):NADP+ oxidoreductase	NADPH [cytoplasm] + 2.0 iron(3+) [extracellular] <=> H+ [cytoplasm] + 2.0 iron(2+) [cytoplasm] + NADP(+) [cytoplasm]	lb
r_4276	Fe(II):NADP+ oxidoreductase	NADPH [cytoplasm] + 2.0 iron(3+) [vacuole] <=> H+ [cytoplasm] + 2.0 iron(2+) [cytoplasm] + NADP(+) [cytoplasm]	lb
r_4292	4-Aminobutyraldehyde:NAD+ oxidoreductase	4-aminobutanal [cytoplasm] + H2O [cytoplasm] + NADP(+) [cytoplasm] <=> gamma-aminobutyrate [cytoplasm] + 2.0 H+ [cytoplasm] + NADPH [cytoplasm]	lb
r_4330	Pyridoxine:NADP+ 4-oxidoreductase	NADP(+) [cytoplasm] + pyridoxine [cytoplasm] <=> H+ [cytoplasm] + NADPH [cytoplasm] + pyridoxal [cytoplasm]	ub
r_4490	Xylitol:NADP+ 4-oxidoreductase (L-xylulose-forming)	H+ [cytoplasm] + NADPH [cytoplasm] + L-xylulose [cytoplasm] <=> NADP(+) [cytoplasm] + xylitol [cytoplasm]	lb
r_4576	(R)-2,3-Dihydroxy-3-methylbutanoate: NADP+ oxidoreductase (isomerizing)	NADP(+) [cytoplasm] + (R)-2,3-dihydroxy-3-methylbutanoate [cytoplasm] <=> H+ [cytoplasm] + NADPH [cytoplasm] + 3-hydroxy-3-methyl-2-oxobutanoate [cytoplasm]	ub
r_4580	(R)-2,3-dihydroxy-3-methylbutanoate: NADP+ oxidoreductase (isomerizing)	NADP(+) [cytoplasm] + (R)-2,3-dihydroxy-3-methylbutanoate [cytoplasm] <=> H+ [cytoplasm] + NADPH [cytoplasm] + (2S)-2-acetolactate [cytoplasm]	ub

APPENDIX 2. Heterologous production pathways (Jouhten *et al*., 2016).

4-Methylthiobutyl-desulfoglucosinolate			
methionine-oxo-acid transaminase	2-oxoglutarate [cytoplasm] + L-methionine [cytoplasm]	<=>	L-glutamate + 4-methylthio-2-oxobutanoic acid [cytoplasm]
methylthioalkylmalate synthase 1	acetyl-CoA + 4-methylthio-2-oxobutanoic acid [cytoplasm] + H2O	<=>	coenzyme A + 2-(2-methylthio)ethylmalic acid [cytoplasm]
2-(2-methylthio)ethylmalate hydroxymutase	2-(2-methylthio)ethylmalic acid [cytoplasm]	<=>	3-(2-methylthio)ethylmalic acid [cytoplasm]
3-(2-methylthio)ethylmalate oxidoreductase (decarboxylating)	3-(2-methylthio)ethylmalic acid [cytoplasm]	<=>	CO2 + 2-Oxo-5-methylthiopentanoic acid [cytoplasm]
methylthioalkylmalate synthase like	acetyl-CoA + 2-Oxo-5-methylthiopentanoic acid [cytoplasm] + H2O	<=>	coenzyme A + 2-(3-methylthio)propylmalic acid [cytoplasm]
2-(3-methylthio)propylmalate hydroxymutase	2-(3-methylthio)propylmalic acid [cytoplasm]	<=>	3-(3-methylthio)propylmalic acid [cytoplasm]
3-(3-methylthio)propylmalate oxidoreductase	3-(3-methylthio)propylmalic acid [cytoplasm]	=>	CO2 + 2-oxo-6-methylthiohexanoic acid [cytoplasm]
2-oxo-acid aminotransferase	L-glutamate + 2-oxo-6-methylthiohexanoic acid [cytoplasm]	<=>	2-oxoglutarate + dihomomethionine [cytoplasm]
dihomomethionine,NADPH:oxygen oxidoreductase (N-hydroxylating, decarboxylating)	2 H(+) + 2 NADPH + 2 oxygen + dihomomethionine [cytoplasm]	=>	CO2 + 2 NADP(+) + 5-methylthiopentanaldoxime [cytoplasm] + 3 H2O
5-methylthiopentanaldoxime,NADPH:oxygen oxidoreductase	H(+) + L-cysteine + NADPH + oxygen + 5-methylthiopentanaldoxime [cytoplasm]	=>	NADP(+) + S-(5-Methylthiopentylthiohydroximoyl)-L-cysteine [cytoplasm] + 2 H2O
S-(5-methylthiopentylthiohydroximoyl)-L-cysteine aminolyase	S-(5-Methylthiopentylthiohydroximoyl)-L-cysteine [cytoplasm] + H2O	=>	NH3 + pyruvate + 5-methylthiopentylthiohydroximate [cytoplasm]
N-hydroxythioamide S-beta-glucosyltransferase	UDP-glucose + 5-methylthiopentylthiohydroximate [cytoplasm]	=>	UDP + 4-methylthiobutyl-desulfoglucosinolate [cytoplasm]
4-methylthiobutyl-desulfoglucosinolate transport	4-methylthiobutyl-desulfoglucosinolate [cytoplasm]	=>	4-methylthiobutyl-desulfoglucosinolate [extracellular]
6-Methylsalicylate			
6-methylsalicylate synthase	acetyl-CoA + H(+) + 3 malonyl-CoA [cytoplasm] + NADPH	=>	3 CO2 + 4 coenzyme A + NADP(+) + 6-methylsalicylate [cytoplasm] + H2O
6-methylsalicylate transport	6-methylsalicylate [cytoplasm]	=>	6-methylsalicylate [extracellular]
6-methylsalicylate exchange	6-methylsalicylate [extracellular]	=>	
8-Epi-cedrol			
epi-cedrol synthase	2-trans,6-trans-farnesyl diphosphate [cytoplasm]	<=>	diphosphate + 8-epi-cedrol [cytoplasm]
8-epi-cedrol transport	8-epi-cedrol [cytoplasm]	=>	8-epi-cedrol [extracellular]
8-epi-cedrol exchange	8-epi-cedrol [extracellular]	=>	
Amorpha-4,11-diene			
amorphadiene synthase	2-trans,6-trans-farnesyl diphosphate [cytoplasm]	<=>	diphosphate + amorpha-4,11-diene [cytoplasm]
amorpha-4,11-diene transport	amorpha-4,11-diene [cytoplasm]	=>	amorpha-4,11-diene [extracellular]
amorpha-4,11-diene exchange	amorpha-4,11-diene [extracellular]	=>	
Apigenin			
tyrosine ammonia lyase	L-tyrosine [cytoplasm]	<=>	NH3 + p-hydroxycinnamic acid [cytoplasm]
4-coumaroyl:CoA ligase	ATP + coenzyme A + p-hydroxycinnamic acid [cytoplasm]	<=>	AMP + diphosphate + 4-coumaroyl-CoA [cytoplasm]
naringenin chalcone synthase	3 malonyl-CoA + 4-coumaroyl-CoA [cytoplasm]	<=>	3 CO2 + 4 coenzyme A + naringenin chalcone [cytoplasm]
naringenin chalcone isomerase	naringenin chalcone [cytoplasm]	<=>	naringenin [cytoplasm]
flavone synthase, apigenin	2-oxoglutarate + oxygen + naringenin [cytoplasm]	<=>	CO2 + succinate + apigenin [cytoplasm] + H2O
apigenin transport	apigenin [cytoplasm]	=>	apigenin [extracellular]
apigenin exchange	apigenin [extracellular]	=>	

Artemisinic acid			
amorphadiene synthase	2-trans,6-trans-farnesyl diphosphate [cytoplasm]	<=>	diphosphate + amorpha-4,11-diene [cytoplasm]
amorpha-4,11-diene monooxygenase1	H(+) + NADPH + oxygen + amorpha-4,11-diene [cytoplasm]	=>	NADP(+) + artemisinic alcohol [cytoplasm] + H2O
amorpha-4,11-diene monooxygenase2	H(+) + NADPH + oxygen + artemisinic alcohol [cytoplasm]	=>	NADP(+) + artemisinic aldehyde [cytoplasm] + 2 H20
amorpha-4,11-diene monooxygenase3	H(+) + NADPH + oxygen + artemisinic aldehyde [cytoplasm]	=>	NADP(+) + artemisinic acid [cytoplasm] + H20
artemisinic acid transport	artemisinic acid [cytoplasm]	=>	artemisinic acid [extracellular]
artemisinic acid exchange	artemisinic acid [extracellular]	=>	
Beta-carotene			
geranylgeranyl diphosphate synthase	isopentenyl diphosphate [cytoplasm] + 2-trans,6-trans-farnesyl diphosphate [cytoplasm]	=>	diphosphate + geranylgeranyl diphosphate [cytoplasm]
all-trans-phytoene synthase	2 geranylgeranyl diphosphate [cytoplasm]	=>	2 diphosphate + all-trans-phytoene [cytoplasm]
all-trans-phytoene monooxygenase	FAD + all-trans-phytoene [cytoplasm]	=>	FADH2 + all-transphytofluene [cytoplasm]
all-trans-phytofluene monooxygenase	FAD + all-transphytofluene [cytoplasm]	=>	FADH2 + zeta-carotene [cytoplasm]
zeta-carotene monooxygenase	FAD + zeta-carotene [cytoplasm]	=>	FADH2 + neurosporene [cytoplasm]
neurosporene monooxygenase	FAD + neurosporene [cytoplasm]	=>	FADH2 + lycopene [cytoplasm]
beta-carotene synthase	lycopene [cytoplasm]	=>	beta-carotene [cytoplasm]
beta-carotene transport	beta-carotene [cytoplasm]	=>	beta-carotene [extracellular]
beta-carotene exchange	beta-carotene [extracellular]	=>	
Butanol			
(R)-3-Hydroxybutanoyl-CoA:NAD+ oxidoreductase	acetoacetyl-CoA [cytoplasm] + H(+) + NADH	<=>	NAD(+) + (R)-3-hydroxybutanoyl-CoA [cytoplasm]
3-hydroxybutyryl-CoA dehydratase	(R)-3-hydroxybutanoyl-CoA [cytoplasm]	<=>	crotonoyl-CoA [cytoplasm] + H2O
trans-2-enoyl-CoA reductase	H(+) + NADH + crotonoyl-CoA [cytoplasm]	<=>	NAD(+) + butanoyl-CoA [cytoplasm]
butanal dehydrogenase	H(+) + NADH + butanoyl-CoA [cytoplasm]	<=>	coenzyme A + NAD(+) + butanal [cytoplasm]
butanol dehydrogenase	H(+) + NADH + butanal [cytoplasm]	<=>	NAD(+) + butanol [cytoplasm]
butanol transport	butanol [cytoplasm]	=>	butanol [extracellular]
butanol exchange	butanol [extracellular]	=>	
Chrysin			
tyrosine ammonia lyase	L-tyrosine [cytoplasm]	<=>	NH3 + p-hydroxycinnamic acid [cytoplasm]
4-coumaroyl:CoA ligase	ATP + coenzyme A + p-hydroxycinnamic acid [cytoplasm]	<=>	AMP + diphosphate + 4-coumaroyl-CoA [cytoplasm]
trans-cinnamate	NADP(+) + 4-coumaroyl-CoA [cytoplasm] + H2O	<=>	H(+) + NADPH + oxygen + cinnamoyl-CoA [cytoplasm]
pinocembrin chalcone synthase	3 malonyl-CoA + cinnamoyl-CoA [cytoplasm]	<=>	3 CO2 + 4 coenzyme A + pinocembrin chalcone [cytoplasm]
pinocembrin chalcone isomerase	pinocembrin chalcone [cytoplasm]	<=>	pinocembrin [cytoplasm]
flavone synthase, chrysin	2-oxoglutarate + oxygen + pinocembrin [cytoplasm]	<=>	CO2 + succinate + chrysin [cytoplasm] + H2O
chrysin transport	chrysin [cytoplasm]	=>	chrysin [extracellular]
chrysin exchange	chrysin [extracellular]	=>	
Cubebol			
cubebol synthase	2-trans,6-trans-farnesyl diphosphate [cytoplasm] + H2O	<=>	diphosphate + cubebol [cytoplasm]
cubebol transport	cubebol [cytoplasm]	=>	cubebol [extracellular]
cubebol exchange	cubebol [extracellular]	=>	

Eriodictyol			
tyrosine ammonia lyase	L-tyrosine [cytoplasm]	<=>	NH_3 + p-hydroxycinnamic acid [cytoplasm]
cinnamate 4-hydroxylase	p-hydroxycinnamic acid [cytoplasm]	<=>	caffeic acid [cytoplasm]
4-coumaroyl:CoA ligase, caffeic acid	ATP + coenzyme A + caffeic acid [cytoplasm]	<=>	AMP + diphosphate + caffeoyl-CoA [cytoplasm]
eriodictyol chalcone synthase	3 malonyl-CoA + caffeoyl-CoA [cytoplasm]	<=>	3 CO_2 + 4 coenzyme A + eriodictyol chalcone [cytoplasm]
eriodictyol chalcone isomerase	eriodictyol chalcone [cytoplasm]	<=>	eriodictyol [cytoplasm]
eriodictyol transport	eriodictyol [cytoplasm]	=>	eriodictyol [extracellular]
eriodictyol exchange	eriodictyol [extracellular]	=>	
Geraniol			
geraniol synthase	geranyl diphosphate [cytoplasm] + H_2O	<=>	diphosphate + geraniol [cytoplasm]
geraniol transport	geraniol [cytoplasm]	=>	geraniol [extracellular]
geraniol exchange	geraniol [extracellular]	=>	
Homoeriodictyol			
tyrosine ammonia lyase	L-tyrosine [cytoplasm]	=>	NH_3 + p-hydroxycinnamic acid [cytoplasm]
cinnamate 4-hydroxylase	p-hydroxycinnamic acid [cytoplasm] + H_2O	=>	caffeic acid [cytoplasm]
S-Adenosyl-L-methionine: 3,4-dihydroxy-trans-cinnamate 3-O-methyltransferase	S-adenosyl-L-methionine + caffeic acid [cytoplasm]	<=>	S-adenosyl-L-homocysteine + ferulic acid [cytoplasm]
Ferulate:CoA ligase (AMP-forming)	ATP + coenzyme A + ferulic acid [cytoplasm]	<=>	AMP + diphosphate + feruloyl-CoA [cytoplasm]
malonyl-CoA:feruloyl-CoA malonyltransferase (cyclizing)	3 malonyl-CoA + feruloyl-CoA [cytoplasm]	<=>	3 CO_2 + 4 coenzyme A + homoeriodictyol chalcone [cytoplasm]
homoeriodictyol chalcone isomerase	homoeriodictyol chalcone [cytoplasm]	<=>	homoeriodictyol [cytoplasm]
homoeriodictyol transport	homoeriodictyol [cytoplasm]	=>	homoeriodictyol [extracellular]
homoeriodictyol exchange	homoeriodictyol [extracellular]	=>	
Hydrocortisone			
corticosterone synthesis 1a	5,7,24(28)-ergostatrienol [cytoplasm] + H(+) + NADPH	<=>	NADP(+) + 5,7-ergostadienol [cytoplasm]
corticosterone synthesis 1b	H(+) + NADPH + 5,7-ergostadienol [cytoplasm]	<=>	NADP (+) + ergosta-5-enol [cytoplasm]
corticosterone synthesis 2 a	H(+) + NADPH + ergosta-5,7,22,24(28)-tetraen-3beta-ol [cytoplasm]	<=>	NADP(+) + ergosta-5-7-22-trien-3-ol [cytoplasm]
corticosterone synthesis 2 b	H(+) + NADPH + ergosta-5-7-22-trien-3-ol [cytoplasm]	<=>	NADP(+) + ergosta-5,22-diene-3-ol [cytoplasm]
corticosterone synthesis 3	3 H(+) + 3 NADPH + 3 oxygen + ergosta-5-enol [cytoplasm] + 6 reduced adrenodoxin [cytoplasm]	=>	3 NADP(+) + pregnenolone [cytoplasm] + 3,4-dimethylpentanal [cytoplasm] + 6 oxidized adrenodoxin [cytoplasm] + 4 H_2O
corticosterone synthesis 4	3 H(+) + 3 NADPH + 3 oxygen + ergosta-5,22-diene-3-ol [cytoplasm] + 6 reduced adrenodoxin [cytoplasm]	=>	3 NADP(+) + pregnenolone [cytoplasm] + 3,4-dimethylpentanal [cytoplasm] + 6 oxidized adrenodoxin [cytoplasm] + $4H_2O$
3,4-dimethylpentanal transport	3,4-dimethylpentanal [cytoplasm]	=>	3,4-dimethylpentanal [extracellular]
3,4-dimethylpentanal exchange	3,4-dimethylpentanal [extracellular]	=>	
corticosterone synthesis 5	H(+) + NADPH + oxygen + pregnenolone [cytoplasm]	=>	NADP(+) + 17-hydroxypregnenolone [cytoplasm] + H_2O
corticosterone synthesis 6	H(+) + NADPH + oxygen + progesterone [cytoplasm]	=>	NADP(+) + 17-hydroxyprogesterone [cytoplasm] + H_2O
corticosterone synthesis 7	NADP(+) + pregnenolone [cytoplasm]	<=>	H(+) + NADPH + progesterone [cytoplasm]
corticosterone synthesis 8	NADP(+) + 17-hydroxypregnenolone [cytoplasm]	<=>	H(+) + NADPH + 17-hydroxyprogesterone [cytoplasm]
corticosterone synthesis 9	H(+) + NADPH + oxygen + 17-hydroxyprogesterone [cytoplasm]	=>	NADP(+) + 11-deoxycortisol [cytoplasm] + H_2O
corticosterone synthesis 10	H(+) + NADPH + oxygen + progesterone [cytoplasm]	=>	NADP(+) + deoxycorticosterone [cytoplasm] + H_2O
corticosterone synthesis 11	H(+) + NADPH + oxygen + reduced adrenodoxin [cytoplasm] + 11-deoxycortisol [cytoplasm]	=>	NADP(+) + oxidized adrenodoxin [cytoplasm] + hydrocortisone [cytoplasm] + H_2O
corticosterone synthesis 12	H(+) + NADPH + oxygen + reduced adrenodoxin [cytoplasm] + deoxycorticosterone [cytoplasm]	=>	NADP(+) + oxidized adrenodoxin [cytoplasm] + corticosterone [cytoplasm] + H_2O
adrenodoxin oxidoreductase	NADPH + 2 oxidized adrenodoxin [cytoplasm]	<=>	H(+) + NADP(+) + 2 reduced adrenodoxin [cytoplasm]

corticosterone transport	corticosterone [cytoplasm]	=>	corticosterone [extracellular]
hydrocortisone transport	hydrocortisone [cytoplasm]	=>	hydrocortisone [extracellular]
corticosterone exchange	corticosterone [extracellular]	=>	
hydrocortisone exchange	hydrocortisone [extracellular]	=>	
Lactate			
lactate dehydrogenase	NADH + pyruvate [cytoplasm]	<=>	NAD(+) + lactate [cytoplasm]
lactate transport	lactate [cytoplasm]	=>	lactate [extracellular]
lactate exchange	lactate [extracellular]	=>	
Luteolin			
tyrosine ammonia lyase	L-tyrosine [cytoplam]	<=>	NH3 + p-hydroxycinnamic acid [cytoplasm]
cinnamate 4-hydroxylase	p-hydroxycinnamic acid [cytoplasm]	<=>	caffeic acid [cytoplasm]
4-coumaroyl:CoA ligase, caffeic acid	ATP + coenzyme A + caffeic acid [cytoplasm]	<=>	AMP + diphosphate + caffeoyl-CoA [cytoplasm]
eriodictyol chalcone synthase	3 malonyl-CoA + caffeoyl-CoA [cytoplasm]	<=>	3 CO2 + 4 coenzyme A + eriodictyol chalcone [cytoplasm]
eriodictyol chalcone isomerase	eriodictyol chalcone [cytoplasm]	<=>	eriodictyol [cytoplasm]
flavone synthase, luteolin	2-oxoglutarate [cytoplasm] + oxygen + eriodictyol [cytoplasm]	=>	CO2 + succinate [cytoplasm] + luteolin [cytoplasm]
luteolin transport	luteolin [cytoplasm]	=>	luteolin [extracellular]
luteolin exchange	luteolin [extracellular]	=>	
Naringenin			
tyrosine ammonia lyase	L-tyrosine [cytoplasm]	<=>	NH3 + p-hydroxycinnamic acid [cytoplasm]
4-coumaroyl:CoA ligase	ATP + coenzyme A + p-hydroxycinnamic acid [cytoplasm]	<=>	AMP + diphosphate + 4-coumaroyl-CoA [cytoplasm]
naringenin chalcone synthase	3 malonyl-CoA [cytoplasm] + 4-coumaroyl-CoA [cytoplasm]	<=>	3 CO2 + 4 coenzyme A + naringenin chalcone [cytoplasm]
naringenin chalcone isomerase	naringenin chalcone [cytoplasm]	<=>	naringenin [cytoplasm]
naringenin transport	naringenin [cytoplasm]	=>	naringenin [extracellular]
naringenin exchange	naringenin [extracellular]	=>	
Nicotianamine			
nicotianamine synthase	S-adenosyl-L-methionine [cytoplasm]	<=>	3 5-methylthioadenosine [cytoplasm] + nicotianamine [cytoplasm]
nicotianamine transport	nicotianamine [cytoplasm]	=>	nicotianamine [extracellular]
nicotianamine exchange	nicotianamine [extracellular]	=>	
5-methylthioadenosine transport	5-methylthioadenosine [cytoplasm]	=>	5-methylthioadenosine [extracellular]
5-methylthioadenosine exchange	5-methylthioadenosine [extracellular]	=>	
p-Hydroxycinnamic acid			
tyrosine ammonia lyase	L-tyrosine [cytoplasm]	<=>	NH3 + p-hydroxycinnamic acid [cytoplasm]
p-hydroxycinnamic acid transport	p-hydroxycinnamic acid [cytoplasm]	=>	p-hydroxycinnamic acid [extracellular]
p-hydroxycinnamic acid exchange	p-hydroxycinnamic acid [extracellular]	=>	
Patchoulol			
patchoulol synthase	2-trans,6-trans-farnesyl diphosphate [cytoplasm] + H2O	<=>	diphosphate + patchoulol [cytoplasm]
patchoulol transport	patchoulol [cytoplasm]	=>	patchoulol [extracellular]
patchoulol exchange	patchoulol [extracellular]	=>	

Pinocembrin			
tyrosine ammonia lyase	L-tyrosine [cytoplasm]	<=>	NH_3 + p-hydroxycinnamic acid [cytoplasm]
4-coumaroyl:CoA ligase	ATP + coenzyme A + p-hydroxycinnamic acid [cytoplasm]	<=>	AMP + diphosphate + 4-coumaroyl-CoA [cytoplasm]
trans-cinnamate	NADP(+) + 4-coumaroyl-CoA [cytoplasm]	<=>	H(+) + NADPH + oxygen + cinnamoyl-CoA [cytoplasm]
pinocembrin chalcone synthase	3 malonyl-CoA + 4-cinnamoyl-CoA [cytoplasm]	=>	3 CO_2 + 4 coenzyme A + pinocembrin chalcone [cytoplasm]
pinocembrin chalcone isomerase	pinocembrin chalcone [cytoplasm]	=>	pinocembrin [cytoplasm]
pinocembrin transport	pinocembrin [cytoplasm]	=>	pinocembrin [extracellular]
pinocembrin exchange	pinocembrin [extracellular]	=>	
Poly-beta-hydroxybutyrate			
(R)-3-Hydroxybutanoyl-CoA:NADP+ oxidoreductase	acetoacetyl-CoA [cytoplasm] + H(+) + NADPH	<=>	NADP(+) + (R)-3-hydroxybutanoyl-CoA [cytoplasm]
poly-beta-hydroxybutyrate polymerase	(R)-3-hydroxybutanoyl-CoA [cytoplasm]	<=>	coenzyme A + poly-beta-hydroxybutyrate [cytoplasm]
poly-beta-hydroxybutyrate transport	poly-beta-hydroxybutyrate [cytoplasm]	=>	poly-beta-hydroxybutyrate [extracellular]
poly-beta-hydroxybutyrate exchange	poly-beta-hydroxybutyrate [extracellular]	=>	
Propane-1,2-diol			
methylglyoxal synthase	glycerone phosphate [cytoplasm]	=>	phosphate + methylglyoxal [cytoplasm]
methylglyoxal reductase	H(+) + NADPH + methylglyoxal [cytoplasm]	<=>	NADP(+) + acetol [cytoplasm]
L-1,2-propanediol dehydrogenase	H(+) + NADH + acetol [cytoplasm]	<=>	NAD(+) + 1,2-propanediol [cytoplasm]
1,2-propanediol transport	1,2-propanediol [cytoplasm]	=>	1,2-propanediol [extracellular]
1,2-propanediol exchange	1,2-propanediol [extracellular]	=>	
Propane-1,3-diol			
glycerol dehydratase	glycerol [cytoplasm]	=>	3-hydroxypropionaldehyde [cytoplasm] + H_2O
1,3-propanediol dehydrogenase	H(+) + NADH + 3-hydroxypropionaldehyde [cytoplasm]	<=>	NAD(+) + 1,3-propanediol [cytoplasm]
1,3-propanediol transport	1,3-propanediol [cytoplasm]	=>	1,3-propanediol [extracellular]
1,3-propanediol exchange	1,3-propanediol [extracellular]	=>	
Resveratrol			
tyrosine ammonia lyase	L-tyrosine [cytoplasm]	<=>	NH_3 + p-hydroxycinnamic acid [cytoplasm]
4-coumaroyl:CoA ligase	ATP + coenzyme A + p-hydroxycinnamic acid [cytoplasm]	<=>	AMP + diphosphate + 4-coumaroyl-CoA [cytoplasm]
resveratrol synthase	3 malonyl-CoA + 4-coumaroyl-CoA [cytoplasm]	<=>	4 CO_2 + 4 coenzyme A + resveratrol [cytoplasm]
resveratrol transport	resveratrol [cytoplasm]	=>	resveratrol [extracellular]
resveratrol exchange	resveratrol [extracellular]	=>	
Taxa-4(20),11(12)-dien-5alpha-acetoxy-10beta-ol			
geranylgeranyl diphosphate synthase	isopentenyl diphosphate [cytoplasm] + 2-trans,6-trans-farnesyl diphosphate [cytoplasm]	=>	diphosphate + geranylgeranyl diphosphate [cytoplasm]
taxa-4(5),11(12)-diene synthase	geranylgeranyl diphosphate [cytoplasm]	=>	diphosphate + taxa-4(5),11(12)-diene [cytoplasm]
taxa-4(5),11(12)-diene monooxygenase	H(+) + NADPH + oxygen + taxa-4(5),11(12)-diene [cytoplasm]	=>	NADP(+) + taxa-4(20),11(12)-dien-5alpha-ol [cytoplasm] + H_2O
taxa-4(20),11(12)-dien-5alpha-ol acetylase	acetyl-CoA + taxa-4(20),11(12)-dien-5alpha-ol [cytoplasm]	=>	coenzyme A + taxa-4(20),11(12)-dien-5alpha-yl acetate [cytoplasm]
taxa-4(20),11(12)-dien-5alpha-yl acetate monooxygenase	H(+) + NADPH + oxygen + taxa-4(20),11(12)-dien-5alpha-yl acetate [cytoplasm]	=>	NADP(+) + taxa-4(20),11(12)-dien-5alphaacetoxy-10beta-ol [cytoplasm] + H_2O
taxa-4(20),11(12)-dien-5alphaacetoxy-10beta-ol transport	taxa-4(20),11(12)-dien-5alphaacetoxy-10beta-ol [cytoplasm]	=>	taxa-4(20),11(12)-dien-5alphaacetoxy-10beta-ol [extracellular]
taxa-4(20),11(12)-dien-5alphaacetoxy-10beta-ol exchange	taxa-4(20),11(12)-dien-5alphaacetoxy-10beta-ol [extracellular]	=>	

Valencene			
trans,trans-farnesyl-diphosphate diphosphate-lyase [cyclizing, (+)-Valencene-forming]	2-trans,6-trans-farnesyl diphosphate [cytoplasm]	<=>	diphosphate + valencene [cytoplasm]
valencene transport	valencene [cytoplasm]	=>	valencene [extracellular]
valencene exchange	valencene [extracellular]	=>	
Vanillin			
3-Dehydroshikimate dehydratase	3-dehydroshikimate [cytoplasm]	=>	protocatechuic acid [cytoplasm] + H2O
aromatic carboxylic acid reductase	ATP + NADPH + protocatechuic acid [cytoplasm]	=>	AMP + NADP(+) + diphosphate + protocatechuic aldehyde [cytoplasm]
O-methyltransferase	S-adenosyl-L-methionine + protocatechuic aldehyde [cytoplasm]	=>	S-adenosyl-L-homocysteine + vanillin [cytoplasm]
vanillin glucosyltransferase	UDP-glucose + vanillin [cytoplasm]	=>	UDP + vanillin glucoside [cytoplasm]
vanillin glucoside transport	vanillin glucoside [cytoplasm] => vanillin glucoside [extracellular]		
vanillin glucoside exchange	vanillin glucoside [extracellular] =>		

APPENDIX 3. Reactions that were removed from possible reaction deletion targets in addition to essential reactions for growth and for the heterologous pathway in question; reactions where there was no flux; transport, exchange and diffusion reactions and reactions without gene annotation. These reactions include reactions e.g. from lipid metabolism.

r_0120	r_2204	r_2297	r_2444	r_2548	r_2628	r_2708	r_2788	r_2940	r_3030	r_3110	r_3190
r_0121	r_2205	r_2298	r_2445	r_2549	r_2629	r_2709	r_2789	r_2941	r_3031	r_3111	r_3191
r_0122	r_2206	r_2299	r_2446	r_2550	r_2630	r_2710	r_2790	r_2942	r_3032	r_3112	r_3192
r_0123	r_2207	r_2300	r_2447	r_2551	r_2631	r_2711	r_2791	r_2943	r_3033	r_3113	r_3193
r_0124	r_2208	r_2368	r_2448	r_2552	r_2632	r_2712	r_2792	r_2944	r_3034	r_3114	r_3194
r_0125	r_2209	r_2369	r_2449	r_2553	r_2633	r_2713	r_2793	r_2945	r_3035	r_3115	r_3195
r_0281	r_2210	r_2370	r_2450	r_2554	r_2634	r_2714	r_2794	r_2946	r_3036	r_3116	r_3196
r_0282	r_2211	r_2371	r_2451	r_2555	r_2635	r_2715	r_2795	r_2947	r_3037	r_3117	r_3197
r_0283	r_2212	r_2372	r_2452	r_2556	r_2636	r_2716	r_2796	r_2948	r_3038	r_3118	r_3198
r_0284	r_2213	r_2373	r_2453	r_2557	r_2637	r_2717	r_2797	r_2949	r_3039	r_3119	r_3199
r_0285	r_2214	r_2374	r_2454	r_2558	r_2638	r_2718	r_2798	r_2950	r_3040	r_3120	r_3200
r_0286	r_2215	r_2375	r_2455	r_2559	r_2639	r_2719	r_2799	r_2951	r_3041	r_3121	r_3201
r_0287	r_2216	r_2376	r_2456	r_2560	r_2640	r_2720	r_2800	r_2952	r_3042	r_3122	r_3202
r_0288	r_2217	r_2377	r_2457	r_2561	r_2641	r_2721	r_2801	r_2953	r_3043	r_3123	r_3203
r_0289	r_2218	r_2378	r_2458	r_2562	r_2642	r_2722	r_2802	r_2954	r_3044	r_3124	r_3204
r_0290	r_2232	r_2379	r_2459	r_2563	r_2643	r_2723	r_2803	r_2955	r_3045	r_3125	r_3205
r_0291	r_2233	r_2380	r_2460	r_2564	r_2644	r_2724	r_2804	r_2956	r_3046	r_3126	r_3206
r_0292	r_2234	r_2381	r_2461	r_2565	r_2645	r_2725	r_2805	r_2957	r_3047	r_3127	r_3207
r_0293	r_2235	r_2382	r_2462	r_2566	r_2646	r_2726	r_2806	r_2958	r_3048	r_3128	r_3208
r_0294	r_2236	r_2383	r_2463	r_2567	r_2647	r_2727	r_2807	r_2959	r_3049	r_3129	r_3209
r_0295	r_2237	r_2384	r_2464	r_2568	r_2648	r_2728	r_2808	r_2960	r_3050	r_3130	r_3210
r_0296	r_2238	r_2385	r_2465	r_2569	r_2649	r_2729	r_2809	r_2961	r_3051	r_3131	r_3211
r_0297	r_2239	r_2386	r_2466	r_2570	r_2650	r_2730	r_2810	r_2962	r_3052	r_3132	r_3212
r_0298	r_2240	r_2387	r_2467	r_2571	r_2651	r_2731	r_2811	r_2963	r_3053	r_3133	r_3213
r_0299	r_2241	r_2388	r_2468	r_2572	r_2652	r_2732	r_2884	r_2964	r_3054	r_3134	r_3214
r_2131	r_2242	r_2389	r_2469	r_2573	r_2653	r_2733	r_2885	r_2965	r_3055	r_3135	r_3215
r_2140	r_2243	r_2390	r_2470	r_2574	r_2654	r_2734	r_2886	r_2966	r_3056	r_3136	r_3216
r_2141	r_2244	r_2391	r_2471	r_2575	r_2655	r_2735	r_2887	r_2967	r_3057	r_3137	r_3217
r_2142	r_2245	r_2392	r_2472	r_2576	r_2656	r_2736	r_2888	r_2968	r_3058	r_3138	r_3218
r_2143	r_2246	r_2393	r_2473	r_2577	r_2657	r_2737	r_2889	r_2969	r_3059	r_3139	r_3219
r_2144	r_2247	r_2394	r_2474	r_2578	r_2658	r_2738	r_2890	r_2970	r_3060	r_3140	r_3220
r_2145	r_2248	r_2395	r_2475	r_2579	r_2659	r_2739	r_2891	r_2971	r_3061	r_3141	r_3221
r_2146	r_2249	r_2396	r_2476	r_2580	r_2660	r_2740	r_2892	r_2972	r_3062	r_3142	r_3222
r_2147	r_2250	r_2397	r_2477	r_2581	r_2661	r_2741	r_2893	r_2973	r_3063	r_3143	r_3223
r_2148	r_2251	r_2398	r_2478	r_2582	r_2662	r_2742	r_2894	r_2974	r_3064	r_3144	r_3264
r_2149	r_2252	r_2399	r_2479	r_2583	r_2663	r_2743	r_2895	r_2975	r_3065	r_3145	r_3265
r_2150	r_2253	r_2400	r_2480	r_2584	r_2664	r_2744	r_2896	r_2976	r_3066	r_3146	r_3266
r_2151	r_2254	r_2401	r_2481	r_2585	r_2665	r_2745	r_2897	r_2977	r_3067	r_3147	r_3267
r_2152	r_2255	r_2402	r_2482	r_2586	r_2666	r_2746	r_2898	r_2978	r_3068	r_3148	r_3268
r_2153	r_2256	r_2403	r_2483	r_2587	r_2667	r_2747	r_2899	r_2979	r_3069	r_3149	r_3269
r_2154	r_2257	r_2404	r_2484	r_2588	r_2668	r_2748	r_2900	r_2980	r_3070	r_3150	r_3270
r_2155	r_2258	r_2405	r_2485	r_2589	r_2669	r_2749	r_2901	r_2981	r_3071	r_3151	r_3271
r_2156	r_2259	r_2406	r_2486	r_2590	r_2670	r_2750	r_2902	r_2982	r_3072	r_3152	r_3272
r_2157	r_2260	r_2407	r_2487	r_2591	r_2671	r_2751	r_2903	r_2983	r_3073	r_3153	r_3273
r_2158	r_2261	r_2408	r_2512	r_2592	r_2672	r_2752	r_2904	r_2984	r_3074	r_3154	r_3274
r_2159	r_2262	r_2409	r_2513	r_2593	r_2673	r_2753	r_2905	r_2985	r_3075	r_3155	r_3275
r_2160	r_2263	r_2410	r_2514	r_2594	r_2674	r_2754	r_2906	r_2986	r_3076	r_3156	r_3276
r_2161	r_2264	r_2411	r_2515	r_2595	r_2675	r_2755	r_2907	r_2987	r_3077	r_3157	r_3277
r_2162	r_2265	r_2412	r_2516	r_2596	r_2676	r_2756	r_2908	r_2988	r_3078	r_3158	r_3278
r_2163	r_2266	r_2413	r_2517	r_2597	r_2677	r_2757	r_2909	r_2989	r_3079	r_3159	r_3279
r_2164	r_2267	r_2414	r_2518	r_2598	r_2678	r_2758	r_2910	r_2990	r_3080	r_3160	r_3280
r_2165	r_2268	r_2415	r_2519	r_2599	r_2679	r_2759	r_2911	r_2991	r_3081	r_3161	r_3281
r_2166	r_2269	r_2416	r_2520	r_2600	r_2680	r_2760	r_2912	r_2992	r_3082	r_3162	r_3282
r_2167	r_2270	r_2417	r_2521	r_2601	r_2681	r_2761	r_2913	r_2993	r_3083	r_3163	r_3283
r_2168	r_2271	r_2418	r_2522	r_2602	r_2682	r_2762	r_2914	r_2994	r_3084	r_3164	r_3284
r_2169	r_2272	r_2419	r_2523	r_2603	r_2683	r_2763	r_2915	r_2995	r_3085	r_3165	r_3285
r_2170	r_2273	r_2420	r_2524	r_2604	r_2684	r_2764	r_2916	r_2996	r_3086	r_3166	r_3286
r_2171	r_2274	r_2421	r_2525	r_2605	r_2685	r_2765	r_2917	r_2997	r_3087	r_3167	r_3287
r_2172	r_2275	r_2422	r_2526	r_2606	r_2686	r_2766	r_2918	r_2998	r_3088	r_3168	r_3288
r_2173	r_2276	r_2423	r_2527	r_2607	r_2687	r_2767	r_2919	r_2999	r_3089	r_3169	r_3289
r_2174	r_2277	r_2424	r_2528	r_2608	r_2688	r_2768	r_2920	r_3000	r_3090	r_3170	r_3290
r_2175	r_2278	r_2425	r_2529	r_2609	r_2689	r_2769	r_2921	r_3001	r_3091	r_3171	r_3291
r_2176	r_2279	r_2426	r_2530	r_2610	r_2690	r_2770	r_2922	r_3002	r_3092	r_3172	r_3292
r_2177	r_2280	r_2427	r_2531	r_2611	r_2691	r_2771	r_2923	r_3003	r_3093	r_3173	r_3293
r_2178	r_2281	r_2428	r_2532	r_2612	r_2692	r_2772	r_2924	r_3004	r_3094	r_3174	r_3294
r_2179	r_2282	r_2429	r_2533	r_2613	r_2693	r_2773	r_2925	r_3005	r_3095	r_3175	r_3295
r_2180	r_2283	r_2430	r_2534	r_2614	r_2694	r_2774	r_2926	r_3006	r_3096	r_3176	r_3296
r_2181	r_2284	r_2431	r_2535	r_2615	r_2695	r_2775	r_2927	r_3007	r_3097	r_3177	r_3297
r_2182	r_2285	r_2432	r_2536	r_2616	r_2696	r_2776	r_2928	r_3008	r_3098	r_3178	r_3298
r_2183	r_2286	r_2433	r_2537	r_2617	r_2697	r_2777	r_2929	r_3009	r_3099	r_3179	r_3299
r_2194	r_2287	r_2434	r_2538	r_2618	r_2698	r_2778	r_2930	r_3010	r_3100	r_3180	r_3300
r_2195	r_2288	r_2435	r_2539	r_2619	r_2699	r_2779	r_2931	r_3011	r_3101	r_3181	r_3301
r_2196	r_2289	r_2436	r_2540	r_2620	r_2700	r_2780	r_2932	r_3022	r_3102	r_3182	r_3302
r_2197	r_2290	r_2437	r_2541	r_2621	r_2701	r_2781	r_2933	r_3023	r_3103	r_3183	r_3303
r_2198	r_2291	r_2438	r_2542	r_2622	r_2702	r_2782	r_2934	r_3024	r_3104	r_3184	r_3304
r_2199	r_2292	r_2439	r_2543	r_2623	r_2703	r_2783	r_2935	r_3025	r_3105	r_3185	r_3305
r_2200	r_2293	r_2440	r_2544	r_2624	r_2704	r_2784	r_2936	r_3026	r_3106	r_3186	r_3306
r_2201	r_2294	r_2441	r_2545	r_2625	r_2705	r_2785	r_2937	r_3027	r_3107	r_3187	r_3307
r_2202	r_2295	r_2442	r_2546	r_2626	r_2706	r_2786	r_2938	r_3028	r_3108	r_3188	r_3308
r_2203	r_2296	r_2443	r_2547	r_2627	r_2707	r_2787	r_2939	r_3029	r_3109	r_3189	r_3309

r_3310	r_3410	r_3490	r_4006	r_3695	r_3789	r_3869	r_4066	r_0263	r_0651	r_2845
r_3311	r_3411	r_3491	r_4007	r_3696	r_3790	r_3870	r_4067	r_0264	r_0652	r_2846
r_3316	r_3412	r_3492	r_4008	r_3697	r_3791	r_3871	r_4068	r_0265	r_0653	r_2847
r_3317	r_3413	r_3493	r_4009	r_3698	r_3792	r_3872	r_4069	r_0266	r_0654	r_2848
r_3318	r_3414	r_3494	r_4010	r_3699	r_3793	r_3873	r_4070	r_0340	r_0655	r_2849
r_3319	r_3415	r_3495	r_4011	r_3700	r_3794	r_3874	r_4071	r_0341	r_0844	r_2850
r_3320	r_3416	r_3496	r_4012	r_3701	r_3795	r_3875	r_4072	r_0342	r_0845	r_2851
r_3321	r_3417	r_3497	r_4013	r_3702	r_3796	r_3876	r_4073	r_0343	r_0847	r_2852
r_3322	r_3418	r_3498	r_4014	r_3703	r_3797	r_3877	r_4074	r_0919	r_0848	r_2853
r_3323	r_3419	r_3499	r_4015	r_3704	r_3798	r_3878	r_4075	r_0920	r_0849	r_2854
r_3324	r_3420	r_3500	r_4016	r_3705	r_3799	r_3879	r_4080	r_0082	r_0850	r_2855
r_3325	r_3421	r_3501	r_4017	r_3706	r_3800	r_3880	r_4081	r_0083	r_1004	r_2856
r_3326	r_3422	r_3502	r_4018	r_3707	r_3801	r_3881	r_4082	r_0357	r_1005	r_2857
r_3327	r_3423	r_3503	r_4019	r_3708	r_3802	r_3882	r_4083	r_0358	r_1006	r_2858
r_3328	r_3424	r_3504	r_4020	r_3709	r_3803	r_3883	r_4084	r_0571	r_1007	r_2859
r_3329	r_3425	r_3505	r_4021	r_3710	r_3804	r_3884	r_4085	r_0100	r_2324	r_2860
r_3330	r_3426	r_3506	r_4022	r_3711	r_3805	r_3885	r_4086	r_0101	r_2325	r_2861
r_3331	r_3427	r_3507	r_4023	r_3712	r_3806	r_3886	r_4087	r_0102	r_2326	r_2862
r_3348	r_3428	r_1449	r_4024	r_3713	r_3807	r_3887	r_4088	r_0103	r_2327	r_2863
r_3349	r_3429	r_1450	r_4025	r_3714	r_3808	r_3888	r_4089	r_0104	r_2328	r_2864
r_3350	r_3430	r_1451	r_4026	r_3715	r_3809	r_3889	r_4090	r_0105	r_2329	r_2865
r_3351	r_3431	r_1452	r_4027	r_3716	r_3810	r_3890	r_4091	r_0106	r_2330	r_2866
r_3352	r_3432	r_1453	r_4028	r_3731	r_3811	r_3891	r_4092	r_0107	r_2331	r_2867
r_3353	r_3433	r_1454	r_4029	r_3732	r_3812	r_3892	r_4093	r_0126	r_2344	r_2868
r_3354	r_3434	r_1455	r_4030	r_3733	r_3813	r_3893	r_4094	r_0127	r_2345	r_2869
r_3355	r_3435	r_1456	r_4031	r_3734	r_3814	r_3894	r_4095	r_0128	r_2346	r_2870
r_3356	r_3436	r_1457	r_4032	r_3735	r_3815	r_3895	r_4098	r_0129	r_2347	r_2871
r_3357	r_3437	r_1458	r_4033	r_3736	r_3816	r_3896	r_4099	r_0130	r_2348	r_2872
r_3358	r_3438	r_1479	r_4034	r_3737	r_3817	r_3897	r_4100	r_0131	r_2349	r_2873
r_3359	r_3439	r_1480	r_4035	r_3738	r_3818	r_3898	r_4101	r_0132	r_2350	r_2874
r_3360	r_3440	r_1481	r_4036	r_3739	r_3819	r_3899	r_4102	r_0133	r_2351	r_2875
r_3361	r_3441	r_1482	r_4037	r_3740	r_3820	r_3900	r_4103	r_0134	r_2360	r_2876
r_3362	r_3442	r_1483	r_3616	r_3741	r_3821	r_3903	r_4104	r_0135	r_2361	r_2877
r_3363	r_3443	r_1484	r_3617	r_3742	r_3822	r_3904	r_4105	r_2308	r_2362	r_2878
r_3364	r_3444	r_1485	r_3618	r_3743	r_3823	r_3905	r_4106	r_2309	r_2363	r_2879
r_3365	r_3445	r_1486	r_3619	r_3744	r_3824	r_3906	r_4107	r_2310	r_2364	r_2880
r_3366	r_3446	r_1487	r_3620	r_3745	r_3825	r_3907	r_4108	r_2311	r_2365	r_2881
r_3367	r_3447	r_1488	r_3621	r_3746	r_3826	r_3908	r_4109	r_2312	r_2366	r_2882
r_3368	r_3448	r_1509	r_3622	r_3747	r_3827	r_3909	r_4110	r_2313	r_2367	r_2883
r_3369	r_3449	r_1510	r_3623	r_3748	r_3828	r_3910	r_4111	r_2314	r_2488	r_3224
r_3370	r_3450	r_1511	r_3624	r_3749	r_3829	r_3911	r_4112	r_2315	r_2489	r_3225
r_3371	r_3451	r_1512	r_3625	r_3750	r_3830	r_3912	r_4113	r_2316	r_2490	r_3226
r_3372	r_3452	r_1513	r_3626	r_3751	r_3831	r_3913	r_4114	r_2317	r_2491	r_3227
r_3373	r_3453	r_1514	r_3627	r_3752	r_3832	r_3914	r_4115	r_2318	r_2492	r_3228
r_3374	r_3454	r_1515	r_3628	r_3753	r_3833	r_3915	r_4116	r_2319	r_2493	r_3229
r_3375	r_3455	r_1516	r_3629	r_3754	r_3834	r_3916	r_4117	r_2320	r_2494	r_3230
r_3376	r_3456	r_1517	r_3630	r_3755	r_3835	r_3917	r_4118	r_2321	r_2495	r_3231
r_3377	r_3457	r_1518	r_3631	r_3756	r_3836	r_3918	r_4119	r_2322	r_2496	r_3232
r_3378	r_3458	r_3963	r_3632	r_3757	r_3837	r_3919	r_4120	r_2323	r_2497	r_3233
r_3379	r_3459	r_3964	r_3633	r_3758	r_3838	r_3920	r_4121	r_2332	r_2498	r_3234
r_3380	r_3460	r_3965	r_3634	r_3759	r_3839	r_3921	r_4122	r_2333	r_2499	r_3235
r_3381	r_3461	r_3966	r_3635	r_3760	r_3840	r_3922	r_4123	r_2334	r_2500	r_3236
r_3382	r_3462	r_3967	r_3636	r_3761	r_3841	r_3923	r_4124	r_2335	r_2501	r_3237
r_3383	r_3463	r_3968	r_3637	r_3762	r_3842	r_3924	r_4125	r_2336	r_2502	r_3238
r_3384	r_3464	r_3969	r_3638	r_3763	r_3843	r_3925	r_4126	r_2337	r_2503	r_3239
r_3385	r_3465	r_3970	r_3639	r_3764	r_3844	r_3926	r_4127	r_2338	r_2504	r_4156
r_3386	r_3466	r_3979	r_3640	r_3765	r_3845	r_3929	r_4128	r_2339	r_2505	r_4203
r_3387	r_3467	r_3980	r_3641	r_3766	r_3846	r_3930	r_4129	r_2340	r_2506	r_4204
r_3388	r_3468	r_3981	r_3642	r_3767	r_3847	r_3931	r_4130	r_2341	r_2507	r_3252
r_3389	r_3469	r_3982	r_3643	r_3768	r_3848	r_3932	r_4131	r_2342	r_2508	r_3253
r_3390	r_3470	r_3983	r_3644	r_3769	r_3849	r_3933	r_4132	r_2343	r_2509	r_3254
r_3391	r_3471	r_3984	r_3645	r_3770	r_3850	r_3934	r_4133	r_3312	r_2510	r_3255
r_3392	r_3472	r_3985	r_3646	r_3771	r_3851	r_3935	r_4134	r_3313	r_2511	r_3256
r_3393	r_3473	r_3986	r_3647	r_3772	r_3852	r_3936	r_4135	r_3314	r_2820	r_3257
r_3394	r_3474	r_3988	r_3672	r_3773	r_3853	r_3941	r_4136	r_3315	r_2821	r_3258
r_3395	r_3475	r_3989	r_3673	r_3774	r_3854	r_3942	r_4137	r_0119	r_2822	r_3259
r_3396	r_3476	r_3990	r_3674	r_3775	r_3855	r_3943	r_4138	r_4280	r_2823	r_3260
r_3397	r_3477	r_3991	r_3675	r_3776	r_3856	r_3944	r_4139	r_4281	r_2824	r_3261
r_3398	r_3478	r_3992	r_3676	r_3777	r_3857	r_3945	r_4140	r_4323	r_2825	r_3262
r_3399	r_3479	r_3993	r_3677	r_3778	r_3858	r_3946	r_4141	r_4332	r_2826	r_3263
r_3400	r_3480	r_3994	r_3685	r_3779	r_3859	r_3947	r_4142	r_0399	r_2827	
r_3401	r_3481	r_3995	r_3686	r_3780	r_3860	r_3948	r_4143	r_0400	r_2836	
r_3402	r_3482	r_3997	r_3687	r_3781	r_3861	r_3949	r_4144	r_0402	r_2837	
r_3403	r_3483	r_3998	r_3688	r_3782	r_3862	r_3950	r_4145	r_0410	r_2838	
r_3404	r_3484	r_3999	r_3689	r_3783	r_3863	r_3951	r_4146	r_0412	r_2839	
r_3405	r_3485	r_4000	r_3690	r_3784	r_3864	r_3952	r_4147	r_0646	r_2840	
r_3406	r_3486	r_4001	r_3691	r_3785	r_3865	r_3953	r_4148	r_0647	r_2841	
r_3407	r_3487	r_4002	r_3692	r_3786	r_3866	r_3954	r_4149	r_0648	r_2842	
r_3408	r_3488	r_4003	r_3693	r_3787	r_3867	r_3955	r_4150	r_0649	r_2843	
r_3409	r_3489	r_4004	r_3694	r_3788	r_3868	r_3956	r_4151	r_0650	r_2844	

APPENDIX 4. Possible selection niche nutrients.

compound name	id for its uptake reaction
2-oxoglutarate	r_1586
acetate	r_1634
ammonium	r_1654
choline	r_1683
D-galactose	r_1710
D-glucose	r_1714
ethanol	r_1761
fumarate	r_1798
gamma-aminobutyrate	r_1800
glycerol	r_1808
L-alanine	r_1873
L-arginine	r_1879
L-asparagine	r_1880
L-aspartate	r_1881
L-aspartate	r_1883
L-glucitol	r_1886
L-glutamate	r_1889
L-glutamine	r_1891
L-glycine	r_1810
L-histidine	r_1893
L-isoleucine	r_1897
L-leucine	r_1899
L-lysine	r_1900
L-methionine	r_1902
L-phenylalanine	r_1903
L-proline	r_1904
L-serine	r_1906
L-threonine	r_1911
L-tryptophan	r_1912
L-tyrosine	r_1913
L-valine	r_1914
ornithine	r_1987
palmitate	r_1993
pyruvate	r_2033
spermidine	r_2051
spermine	r_2052
succinate	r_2056