In Silico Design for Systems-Based Metabolic Engineering In *Escherichia Coli* for The Bioconversion of Aerobic Glycerol to Succinic Acid

Albert Enrique Tafur Rangel
Universidad de Los Andes https://orcid.org/0000-0002-9428-183X
Wendy Lorena Rios Guzman
Universidad de Los Andes
Carmen Elvira Ojeda Cuella
Universidad de Los Andes
Daissy Esther Mejia Perez
Universidad de Los Andes
Ross Carlson
Montana State University Bozeman
Jorge Mario Gomez Ramirez
Universidad de Los Andes
Andres Fernando Gonzalez Barrios (andgonza@uniandes.edu.co)
Universidad de Los Andes https://orcid.org/0000-0003-2517-2020

Research

Keywords: Systems metabolic engineering, transcriptomics, genome-scale model, Escherichia coli, metabolic modeling

Posted Date: December 1st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-113854/v1

Abstract

Background

Glycerol has become an interesting carbon source for industrial processes as consequence of the biodiesel business growth since it has shown promising results in terms of biomass/substrate yields. Selecting the appropriate metabolic targets to build efficient cell factories and maximize the desired chemical production in as little time as possible is a major challenge in industrial biotechnology. The engineering of microbial metabolism following rational design has been widely studied. However, it is a cost-, time-, and laborious-intensive process because of the cell network complexity; thus, to be proficient is needed known in advance the effects of gene deletions.

Results

An *in silico* experiment was performed to model and understand the effects of metabolic engineering over the metabolism by transcriptomics data integration. In this study, systems-based metabolic engineering to predict the metabolic engineering targets was used in order to increase the bioconversion of glycerol to succinic acid by *Escherichia coli*. Transcriptomics analysis suggest insights of how increase the glycerol utilization of the cell for further design efficient cell factories. Three models were used; an *E. coli* core model, a model obtained after the integration of transcriptomics data obtained from *E. coli* growing in an optimized culture media, and a third one obtained after integration of transcriptomics data obtained from *E. coli* after adaptive laboratory evolution experiments. A total of 2402 strains were obtained from these three models. Fumarase and pyruvate dehydrogenase were frequently predicted in all the models, suggesting that these reactions are essential to increasing succinic acid production from glycerol. Finally, using flux balance analysis results for all the mutants predicted, a machine learning method was developed to predict new mutants as well as to propose optimal metabolic engineering targets and mutants based on the measurement of importance of each knockout's (feature's) contribution.

Conclusions

The combination of transcriptome, systems metabolic modeling, and machine learning analyses revealed versatile molecular mechanisms involved in the utilization of glycerol. These data provide a platform to improve the prediction of metabolic engineering targets to design efficient cell factories. Our results may also work a guide platform for the selection/engineering of microorganisms for production of interesting chemical compounds.

Background

Shifting from petrochemical sources to renewable, abundant, and inexpensive feedstocks to obtain valuable chemicals has become a promising goal for the chemical industry [1]. Biodiesel industry has increased in the last years by using renewable raw materials, but it generates large amounts of glycerol becoming it a burden. The bioconversion of glycerol is a potential route to increasing the use bio-based succinic acid industry, a critical building-block chemical with an attractive market. The availability of three pathways for succinic acid production (Fig. 1) [2], the adaptability to different environments, and the accessibility of metabolic engineering and omics tools make *Escherichia coli* an attractive cell factory. However, some challenges, such as low growth rate and yield, the use of rich medium, the generation of by-products, and various anaerobic requirements, need to be overcome

for bio-based succinic acid production, considering cost-effective issues compared to the petroleum-based approach.

The main goal of using microbial cell factories is design cheap and high-yield biotechnology-based conversion processes. A significant problem to be solved is how to enhance cell growth while using its capabilities to obtain a high yield target chemical product. A classical approach for that is adaptive laboratory evolution – ALE, which is based on the selection of microorganisms with superior production capability after random mutagenesis screening. Another approach to strain improvement is metabolic engineering, which uses genetic manipulation to optimize the production of desired compounds. Metabolic engineering select targets that increases the productivity based on the rationality of trial-and error development cycles and an understanding of the routes playing a significant role in the synthesis. Strain design with this method has been extensively applied to use and/or produce interesting compounds [2–6], including bio-based organic acids by substrate transport enhancement, gene overexpression, and deletion [6–11]. However, making the strain industrially competitive requires much time, effort, and high cost [12].

Once DNA was discovered in the last century, a new approach called metabolic network modeling for the study of cellular metabolism was developed [13]. It allows to determine how several pathways in a cell can interact, as well as to elucidate basic microbial processes [14]. The first genome-scale metabolic network was described in 1999 and in 2002 was reported the use of metabolic modeling to analyze recombinant pathways [15]. Ever since, several models have been developed with significant accuracy and useful predictions [16] that can be used to guide experimental studies [13, 17]. COnstraints-Based Reconstruction and Analysis – COBRA - methods make possible to predict, given a cellular objective function, attractive targets to increase or maximize biochemical yields, and to determine perturbations after genetic manipulations of the cell [18, 19]. OptKnock, OptStrain, OptForce, and OptReg are some COBRA methods developed to predict metabolic engineering targets for cell optimization by using Gene-Protein-Reaction -GPR- relationship [17, 20–23].

OptKnock applies a Flux Balance Analysis -FBA- approach for simulating genome-scale metabolic models. It assumes that each organism's metabolic network has been tuned through evolution for some objective function, be it a maximal growth rate or energy efficiency (e.g., minimal ATP utilization). While this assumption may be valid for wild-type (WT) organisms that have evolved over many hundreds or thousands of generations, it may be less appropriate for engineered mutants (KO) because they have been engineered in a controlled environment and unexposed to the same evolutionary forces. Hypothesizing that mutant organisms are unable to immediately adapt their metabolic network to achieve the WT objective function, computational tools such as Minimization of metabolic adjustment (MOMA) was developed [24]. This approach is mathematically formalized as a quadratic programming (QP) problem, finding a suboptimal flux profile that is a minimal Euclidean distance from the WT (WT-FBA) and the genetically perturbed (KO-FBA) organisms. FBA combined with MOMA evaluation after OptKnock prediction could provide more accurate prediction of the immediate metabolic response to KO than FBA does on its own. However, there could exists a large list of knockouts combinations that could be obtained when computational tools are used and select which test in lab can be laborious.

Several approaches to optimize cell factories have been developed, but conventional and computational approaches have not always been successful due to unexpected changes in the cell where exits an intracellular complex interconnected network of genes, proteins, and reactions. Systems metabolic engineering has emerged as an approach that integrates metabolic engineering and combined metabolic and 'omics' network models. This could be beneficial for genome-scale modeling because reduce the solution space and generates accurate

predictions [12, 25–28]. Mainly considering that under certain environmental conditions there are a limited number of reactions which are active according to transcriptional responses and other regulation phenomena to provide beneficial improvements for the cell bioconversion process [29–35].

In this study, systems metabolic engineering for overproduction of succinic acid from glycerol in *E. coli* was used through the integration of transcriptomics data to metabolic models and classification trees analysis using random forest to classify gene targets predicted by OptKnock. Our strategy took advantage of transcriptomics data obtained from an evolved *E. coli* in glycerol, and an optimized culture media. These data were subsequently integrated into a metabolic network models to predict targets using OptKnock. Predicted combinations were then evaluated using FBA, flux variability analysis, and MOMA to determine the effects of gene-reaction knockout in the cell. Finally, predicted target reactions were evaluated using random forest to determine the importance of each target using the succinic acid production, growth rate, and Euclidian distance between the WT strain and each mutant as response variables.

Results

Transcriptional response of E. coli for aerobic glycerol consumption

A cell is considered as a complex system where a large number of processes are carried out. These processes then involve an interaction between, genes, transcripts, proteins, metabolites, reaction, among others [12, 36, 37]. Metabolic models are reconstructed by using the genome information; however, it is well known that metabolism is given by environmental conditions by passing through a cell regulation process. This causes that some genes can be turned on and off under certain conditions. To determine which reactions are active to obtain high accurate models two transcriptomic profile was obtained from an ALE experiment and an optimized culture medium.

Differentially Expressed Genes (DEGs) were determined using the deseq2 statistical package after filtering out low count reads with an average value of < 100. Significantly DEGs were defined as those whose abundance had at least a log2-fold change [(log2 FC) >|2|] between the reference condition (glycerol-based medium) and a chosen experimental condition (optimized culture medium and evolved strain) at a false discovery rate (FDR)-corrected P value of < 0.05. Relevant genes with log2 FC >|1| for glycerol metabolism or under the same regulon were taken into account. Figure 2 shows the distribution of DEGs using a Log2 FC ≥|2| for one strain growing in the optimized culture medium and one evolved strain growing in the same optimized medium. This analysis determined that 478 genes were differentially expressed, with 222 genes down-regulated and 256 up-regulated for the optimized medium, and 431 DEGs for the evolved strain, which 223 genes were down-regulated and 208 genes were up-regulated. When comparing DEGs in the optimized medium and those in the evolved strain, 59 down-regulated genes were found to be unique in the evolved strain and 58 unique genes for optimized medium. In this context, 47 and 95 up-regulated genes were found to be unique in the evolved strain and the optimized medium, respectively (Fig. 2).

DEGs were classified into the following three groups using GO analysis: biological processes, molecular functions, and cellular components. The shared down-regulated genes predominantly included those involved in metabolic process (cellular, organic substances, nitrogen compounds, and primary metabolic processes), chemicals, stress and stimulus responses, and heterocyclic compound systems. Between down-regulated genes, we found *phoB* and *phoR*, which are involved in phosphorous uptake and metabolism since, under excess phosphorous, PhoR

inactivates *phoB* [38]. Figure 3 shows the level 2 GO terms for unique down- and up-regulated genes in both conditions using blast2GO [39]. The 117 unique down-regulated genes at Log2 FC ≥|2| and an adjusted p-value of ≤0.05 were classified into 15 functional groups. Two GO terms, signaling and locomotion, were only present for the evolved strain, and one GO term, multi-organism processes, was only present for the optimized culture condition in down-regulated genes (Fig. 3A).

GO analysis revealed that shared up-regulated DEGs (Fig. 2) are involved mostly in the metabolic process (51%), including GO terms such as cellular, organic substances, primary, and nitrogen compounds process. 11% of up-regulated genes were associated with biosynthetic processes and the establishment of localization. The main GO terms for molecular functions were those involved in binding activity (66%), counting ions, heterocyclic compounds, organic cyclic compounds, small molecules, and protein binding, followed by transferase activity (10%) and transmembrane transporter activity (9%). About 42% of the DEGs categorized in cellular functions were implicated in membrane GO terms, with 17% in the cell periphery and 16% in the cytoplasm.

Glycerol metabolism in *E. coli* is mediated by *glp* operons. In consequence, transcriptomic analysis shows shared up-regulation of *glpBCFKQTX* genes. The changes in bacterial gene expression in response to glycerol utilization are summarized in Table 1. During glycerol utilization, GlpF permease facilitates glycerol entry into *E. coli* for further transformation into glycerol-3-phosphate (Gly-3-P) by GlpK under aerobic conditions. Comparing *glpK* expression with the values obtained for other genes in the glp regulon showed that *glpK* was one of the most highly expressed genes. However, a difference of ~ 1 log2 FC between the evolved strain and the optimized culture condition was exhibited in the *glpFKX* operon (Table 1). As a consequence of the regulatory network, an increase in the expression of *glpX* was detected (2.76 Log2 FC), which is part of the *glpFKX* operon, and works as an alternative fructose-1,6-bisphosphatase involved in gluconeogenesis by catalyzing the hydrolysis of fructose 1,6-bisphosphate to fructose 6-phosphate [40]. Overexpression of *glpX* has been shown to increase hydrogen production [41]. Additionally, transcriptomic analysis showed up-regulation of both flavin oxidases, *glpD* and *glpABC.*

Table 1
Differential expression of genes involved in glycerol metabolism.

Refseq tag (ECOLC_RS)	Gene name	Old locus tag	Product	Log2FC Exp 3	Log2FC Evolved
01540	*glpD*	EcolC_0288	aerobic glycerol-3-phosphate dehydrogenase	1.69	1.21
07540	*glpC*	EcolC_1408	anaerobic glycerol-3-phosphate dehydrogenase subunit C	2.61	3.59
07545	*glpB*	EcolC_1409	anaerobic glycerol-3-phosphate dehydrogenase subunit B	2.87	3.74
07550	*glpA*	EcolC_1410	sn-glycerol-3-phosphate dehydrogenase subunit A	1.28	2.00
07555	*glpT*	EcolC_1411	glycerol-3-phosphate transporter	5.41	4.36
07560	*glpQ*	EcolC_1412	glycerophosphoryl diester phosphodiesterase	5.25	5.52
22045	*glpF*	EcolC_4091	aquaporin	4.17	3.03
22050	*glpK*	EcolC_4092	glycerol kinase	5.36	4.37
22055	*glpX*	EcolC_4093	fructose 1,6-bisphosphatase	2.76	2.05
10840	*fumA*	EcolC_2018	fumarate hydratase	1.46	-0.60
20740	*frdA*	EcolC_3856	fumarate reductase flavoprotein subunit	0.68	1.71
20745	*frdB*	EcolC_3857	fumarate reductase iron-sulfur subunit	0.89	1.87
20750	*frdC*	EcolC_3858	fumarate reductase subunit C	0.74	1.72
20755	*frdD*	EcolC_3859	fumarate reductase subunit D	0.16	1.09

The electron-transport chains of *E. coli* are composed of many different dehydrogenases and terminal reductases. Glycerol metabolism in *E. coli* uses oxygen as the main electron acceptor but it could employ also fumarate under anaerobic conditions by encoding a fumarate reductase complex under anaerobic conditions [42, 43]. Table 1 shows Log2 FC for *fumA* and *frdABCD* genes in *E. coli*. The *fumA* gene was encoded for abundant fumarase, predominantly expressed in the optimized culture medium (1.55 Log2 FC), but not for the evolved strain (−0.53 Log2 FC). FumA have been reported to be predominant expressed under aerobic conditions [44]. Under aerobic conditions, the catalysis of succinate to fumarate interconversion are mediated by the succinate dehydrogenase complex encoded by *sdhABCD* [43]. However, in this study *sdhABCD* genes were not found to be differentially expressed in any of the culture conditions. Interestingly, among upregulated genes in the adapted strain, a difference of ~ 1 log2 FC in the expression of the fumarate reductase genes (*frdABCD*), which is used in anaerobic growth, was observed over the optimized culture condition.

The maltose operon of *E. coli* consists of genes that encode proteins involved in the uptake and metabolism of maltose and maltodextrins. These genes have been found to be highly associated with up-regulation under glycerol utilization as a carbon source, and changes in the level of *glpK* transcription had a significant effect on *malT* transcription [45]. In this study, *malEFKMTPQ* genes were shown to be up regulated in both conditions. For *malT*, the log2 FC was more highly expressed in the optimized culture condition than in the evolved strain. The

same behavior was observed for *glpK*. Thus, high expression of this regulon in this study could be presumably linked to the high expression of the *glpK* gene since they showed similar log2 FC.

As a result of glycerol metabolism, acetate is mainly generated. In our analysis, the phosphate acetyltransferase encoded by *pta*, which catalyzes the reversible conversion between acetyl-CoA and acetylphosphate [46, 47], was found to be up regulated (~ 2.30 Log2 FC). Also, the *atpABCDEFGH* genes have a role in the generation of ATP from ADP and phosphate. These genes were observed to be up-regulated, with similar log2 FC, except for *atpA*, which had a difference of around 1 log2 FC in the optimized culture medium with respect to the evolved strain.

Predicting potential metabolic engineering targets for succinic acid overproduction

Genome-scale metabolic (GEMs) models are defined as a complete set of reactions involved in cell metabolism, given by genome annotation, regardless of whether the annotated metabolic genes are expressed in a given environment. This assumption could be correct in genome-scale models because core models represent the central metabolism, but the full potential of GEMs remains largely unexploited [48]. To avoid this situation and to evaluate the effects of use a core or a large model to predict metabolic engineering targets three models were used: a core model (ECC2) and two models obtained after the integration of transcriptomics data that can help to elucidate the actual state of the metabolic network *in vivo* for further metabolic engineering.

Metabolic model reconstruction and transcriptomics integration

For the integration process, a reconstruction of the metabolic model for *E. coli* ATCC 8739 was carried out based on the iEcolC 1368 [49], iEC1349_Crooks [50], and iML1515 models [51]. Extensive manual curation was conducted to fill pathway gaps. Transport and exchange reactions were added or eliminated, enabling nutrient uptake and by-product secretions. Finally, the resulting model was designated iTA1338, and it involved 2032 metabolites, 2804 reactions, and 1338 genes (Supplementary File S1). After that, using GIMME, it was constructed context-specific metabolic networks departing from the iTA1338 model for two types of strains: (1) WT *E. coli* ATCC 8739 growing in an optimized culture medium (iTA818) (Supplementary file S1) and (2) *E. coli* ATCC 8739 strains evolved to grow on glycerol (iTA821) (Supplementary file S1), manual curation was carried for iTA821 model based on GapFind and GapFill results.

Figure 4 illustrates the number of reactions obtained for each model after transcriptomic integration. The same growth rate was observed after integration; however, flux distribution in 24 reactions was exhibited (Fig. 4B). The reactions only present in iTA821 are mainly associated with the transport inner membrane (14). Other unique reactions in iTA821 were mapped to be linked to the citric acid cycle, cofactor and prosthetic group biosynthesis, glutamate metabolism, inorganic ion transport and metabolism, the nucleotide salvage pathway, oxidative phosphorylation, and pyruvate metabolism, among others. Unique reactions of iTA818 were mainly associated with transport, including the transport outer membrane porin (218), transport inner membrane (50), and transport outer membrane (15), followed by cell envelope biosynthesis (37), the nucleotide salvage pathway (24), glycerophospholipid metabolism (14), alternate carbon metabolism (12), and cofactor and prosthetic group biosynthesis (7), among others.

In silico systems metabolic engineering targets prediction

To design *E. coli* strains that overproduce succinic acid from glycerol, OptKnock was used [22]. Before predicting the reaction target to overproduce succinic acid, both metabolic networks were pre-processed. The goal of pre-processing was to obtain a smaller set of selected reactions that could serve as valid targets for gene knockouts. First, all reactions displaying maximum and minimum fluxes equal to zero were removed from the set of potential reactions to be knocked out. Next, all reactions that had been experimentally found to be essential for growth were removed from consideration [52]. Also, the reactions that were found to be computationally essential were not considered, as well as non-gene-associated reactions.

Ten OptKnock rounds of mutant prediction were carried out. In each round, the set of reactions was set up to 1, 2, 3, … 10, and 100 mutants were requested per round. ECC2, iTA818, and iTA821 models were used to predict mutants of succinic acid overproducers. 811, 806, and 785 possible mutants were obtained from ECC2, iTA818, and iTA821 model, respectively. Figure 5 describes the frequency of the reactions predicted in all the possible mutants. It can be seen that 30 reactions were above the average of the frequency. Reactions acetate kinase (ACKr), fructose 6-phosphate aldolase (F6PA), fumarase (FUM), pyruvate dehydrogenase (PDH), pyruvate formate lyase (PFL), phosphotransacetylase (PTAr), succinate dehydrogenase (SUCDi), triose-phosphate isomerase (TPI), glycerol-3-phosphate dehydrogenase-NADP (G3PD2), and glycerol dehydrogenase (GLYCDx) were frequently predicted for all models. It is important to mention that G6PDH2r, LDH_D, PGL, and POX were not predicted to be part of models iTA818 and iTA821 after integration.

Interestingly, in the complete set of reactions predicted, PDH was the most frequent target reaction, followed by FUM in all models (Fig. 5) and minimal variations in the knockout frequency were observed for these reactions. Figure 5A shows the plot of the first two principal components of the principal components analysis (PCA), representing the variability of 89% of the data. This analysis shows how PDH and FUM knockouts are closely related to succinic acid overproduction from glycerol. Regions of high variability are clustered along with the first principal component, presenting a value of zero for the first principal component. This indicates that the factors that make up the first principal component are critical for high titers. The contributions of different models to the first two principal components of the PCA are shown in Fig. 5B, and they are indicative of the relative influence on the variability in knockout predictions given by transcriptomic integration.

A cluster analysis between the reaction frequency for each *k* deletion showed that elimination of acetate, formate, and lactate by-products mediated by POX, PFL, and LDH_D are highly related to PDH and FUM deletion (Fig. 6). This phenomenon, probably due to PDH deletion results in reduced conversion of pyruvate to acetyl-CoA, which is the main substrate in ACKr and PTAr reaction to generate acetate (Fig. 1), a competitive by-product on succinate production [46]. Then, if PDH deletion is not carried out, ACKr and PTAr knockouts would become essential to increasing succinate production, as well as minimizing costs in the separation process [53, 54].

Since metabolic manipulation of a cell results in a stress process, negative impact of deletions on the maximum growth rate can be observed. To determine the effects of reaction knockouts over the cell, FBA was carried out and Euclidean distance was calculated for each mutant predicted. Figure 7 illustrates the relationship between the number of knockouts, succinic acid production, and growth rate using FBA and the Euclidian distance between the WT and mutant strains using MOMA. It can be seen that the number of reactions knocked is highly related to high succinic acid production rates due to the elimination of competitive by-products, such as acetate, formate, and lactate, requiring at least 3–4 deletions. The highest succinate production (~ 8.5 mmol/gDW/h) was observed in mutants predicted in ECC2 when 9 or 10 reactions were deleted. However, this implies a substantial reduction in the growth rate to ~ 4% compared to the WT strain. Thus, selecting these mutants is unrealistic for the industrial

production of succinic acid. The same behavior in the reduction of the growth rate was observed for those mutants that required more than 6 deletions in mutants predicted in iTA818 and iTA821. In contrast, a considerable reduction in the growth rate (28% of the WT growth rate) as well as an increasing succinic acid production rate (around 30% more than those with 9–10 knockouts) for those mutants with 6 knockouts was observed. In addition, it was observed that there is no direct correlation, in the same magnitude for all the mutants, between the Euclidean distance and the numbers of knockouts in each mutant. However, Fig. 7 shows amplifications in the Euclidian distance between the WT and the mutants when succinate production and knockout numbers increases and growth rate decreases.

Identification of critical metabolic targets and potential mutants

OptKnock results are a large list of a knockouts combinations where maximum product synthesis occurs at maximum growth rate reachable [22]. However, it has been observed that the optimal solution of the target given by OptKnock is not necessarily growth coupled and some mutants predicted does not increase the product target. In consequence, select a mutant to be tested in the lab could be really difficult and probably result in a laborious process. Assuming that each mutant product growth "coupled" predicted will result in a successful biological production, this mutants can ensure high productivity over time and initially solve this situation [55]. To identify growth-coupled production solutions, a Cobra Toolbox function was used verifying the minimum and maximum production rate given a set of reactions to be knocked. As a result, when the maximum growth rate is achieved the same minimum and maximum flux for the desired product should be obtained. 1799 mutants were predicted to be growth coupled, 539 to be growth-coupled non-unique (maximum flux - minimum flux > 0.1), and 64 mutants were categorized as not growth-coupled (maximum flux < 0.1). For the mutants categorized as growth-coupled non-unique, an FBA was carried out to predict the succinic acid production rate (Fig. 7), where 279 mutants were predicted to have a difference between the maximum production rate predicted by the function and FBA < 2, resulting in 2078 *in silico* mutants that overproduce succinic acid.

In order to filter and select potential mutants to be tested in the lab, a random forest model to predict the importance of each reaction knockout was developed based on the OptKnock predictions. Each possible combination of reactions using binary values that increase the succinic acid production was associated with the flux of the extracellular succinic acid and biomass reaction obtained by FBA and the Euclidian distance obtained by MOMA. The data set was divided into two groups, 70% for training and 30% for the test. Following feature selection and cross validation, a robust model that associated any combination of 58 reaction variables to a predicted growth rate and succinic acid production ratio was obtained. A measure of importance the contribution of each feature to the random forest model is shown in Fig. 8 indicated by *IncNodePurity*. This model exhibited a mean square error (MSE) value when using the reaction flux of EX_succ_e flux obtained by FBA as variable response of 0.293. For the growth rate (biomass reaction) as a response variable, the MSE value was 0.0002. Finally, when the Euclidian distance for each mutant was used as the response variable, the MSE value was 9175.158, indicating that the Euclidian distance is not a good response variable to predict cell behavior when using the random forest model. Moreover, this result allows use machine learning models to predict a largest number of mutants than those obtained by OptKnock in terms of growth rate and succinic acid production since OptKnock is high time consuming.

Figure 8A shows that PFL, LDH_D, GLYCDx, G3PD2, PDH, and POX are the most important reactions to increase the amount of succinic acid. These reactions are mainly associated with the GldA–DhaKLM fermentative route and the Gly-3-P route (Fig. 1) in glycerol utilization [46], as well as acetyl-CoA generation given by the PDH knockout. In

around 24% of the mutants predicted, a combination of GLYCDx and G3PD2 reactions was found to increase the succinic acid production. However, POX and LDH_D reactions were not present in iTA818 and iTA821 models, and PDH, G3PD2, and PFL were also found to be the most important reactions, predicted to have an effect over the growth rate (Fig. 8B).

The pyruvate dehydrogenase complex is a critical connection point between glycolysis and the TCA cycle, both of which function during aerobic respiration through catalyzing the conversion of pyruvate to acetyl coenzyme A (acetyl-CoA) [56]. PDH deactivation results in PFL carrying the flux from pyruvate to acetyl-CoA [57]. Simple reaction knockouts show that PDH deletion results in a growth rate reduction of ~ 5%. Additionally, 5 reactions (FUM, GAPD, PGK, PGM, and TPI) were predicted to have the most significant reduction (8–10%) in the growth rate during glycerol utilization. Of those reactions, only FUM has a significant positive effect over the succinate production when this deletion is carried out alone. However, in mutants in which both FUM and PDH were predicted (59.45%), TPI appeared in around 12.60% (Fig. 5). Then, the deletion of genes associated with TPI in addition to FUM and PDH reactions could negatively affect the growth rate.

Discussion

Glycerol metabolism in *E. coli* have been described in the literature. However, cell changes are carried out as response to stress situations. In this study, two conditions were tested for transcription response in *E. coli* for further integration to metabolic network modeling. Gene expression-wide analyses reveal how cell have the ability to avoid glycerol toxicity increasing the consumption. The most striking response to glycerol consumption and the possible mechanism to optimize succinic acid production from glycerol was revealed by the combination of transcriptome, metabolic modeling, and machine learning analyses.

After glycerol incorporation in the cell mediated by GlpF, glycerol can be metabolized through two pathways. The first is mediated by the glycerol kinase GlpK through phosphorylation of glycerol to Gly-3-P, followed by GlpD activity under aerobic conditions, leading to dihydroxyacetone phosphate -DHAP- (Fig. 1). The alternative pathway consists of an oxidation step by glycerol dehydrogenase (GldA) to yield dihydroxyacetone (DHA), followed by phosphorylation by DHA kinase (DhaK) to give DHAP, as well. In this study, over expression of *glpK* was observed in both conditions, with a difference of around 20%. This result is not surprising since the GlpK-mediated reaction is the rate-limiting step in glycerol utilization [58]. However, it has been observed that under the optimized culture conditions the glycerol utilization rate is higher than the evolved conditions, suggesting that should exits other mechanism in the cell to enhance glycerol utilization. Gly-3-P is the first intermediate between the glycerol pathway and the TCA cycle, as well as the biosynthesis and catabolism of lipids; however, accumulation of Gly-3-P can become toxic. Thus, it is carefully regulated [40]. The export of Gly-3-P could be mediated by *phoE* and *ompF* membrane porins; however, down-regulation of *phoE* (−8.67 and − 9.04 Log2 FC for the optimized culture and the evolved strain, respectively) and up-regulation of *ompF* (Log2 FC 2.43) in the optimized culture suggest that it could play an essential role in *E. coli* ATCC 8739 glycerol metabolism at high uptake rates avoiding toxicity.

The marked up-regulation of *glpQ* (5.351 and 5.597 Log2 FC for the optimized culture and evolved strain, respectively), which catalyzes the hydrolysis of glycerol-phosphodiesters to an alcohol plus Gly-3-P together with *ompF* could explain the partially higher transcript abundance of *glpT* since the externally generated (or supplied) Gly-3-P activates GlpT [59, 60]. This protein exchanges Gly-3-P for phosphate, avoiding the toxicity of both Gly-3-P and the inorganic phosphates [40]. As result, and considering that phosphate is necessary to increase glycerol utilization, autoregulation of the PhoB/PhoR two-component regulatory system need to be down expressed. Down

regulation of PhoB/PhoR was observed in this study which could explain the achievement of optimal density [61], as well as contribute to the regulation of glycerol phosphate metabolism [62].

The transcriptional analysis also identified the differential expression of both flavin oxidases, *glpD* and *glpABC.* Once Gly-3-P is in the cytoplasm, it is oxidized to dihydroxyacetone phosphate by one of two flavin-dependent oxidases encoded by *glpD* or *glpABC* genes under aerobic or anaerobic conditions, respectively [40, 46]. In the presence of oxygen or nitrates, GlpD transfers electrons to the respective terminal oxidizes. In contrast, under anaerobic conditions the GlpABC system transfers the electrons to fumarate or nitrates [63]. GlpD up-regulation was expected since culture conditions were under aerobic conditions, but higher expression of the *glpABC* system was surprising. Over expression of *glpABC* under aerobic conditions could be elucidated because of the activation of fumarate reductase enzymes (Table 1) in the evolved strain as results of high cell densities during the ALE process. However, in glycerol fermentation studies, the Δ*frdA* mutant has been shown to be beneficial for glycerol fermentation because it prevents the negative impact of hydrogen by maintaining suitable redox conditions [64]. Moreover, its activity could be supported by *sdhABCD* since they are structurally and functionally homologous [65]. Therefore, we hypothesized that *frdABCD* up-regulation could be the reason why enhancement in the glycerol utilization was not observed in the evolved strain, even when an optimized culture medium was employed.

The insights on the molecular adaptive responses of *E. coli* to glycerol consumption revealed by the transcriptional datasets identified a marked *hdeAB* up-regulation only in the evolved strain. This is attractive since HdeAB are periplasmic proteins that play a role in optimal protection at low pH [66, 67]. Therefore, differences in *hdeAB* up-regulation in the evolved strain and the optimized culture medium probably occur because acetate is the main product in glycerol utilization and under ALE conditions the pH was not controlled. Moreover, the addition of a phosphate buffer system using the salts Na_2HPO_4 and KH_2PO_4 provides the culture medium used directly for the optimized condition with a buffering capacity (Ng, 2018).

It was observed that the main and preferable route for glycerol consumption is the pathway mediated by GlpK since this gene was highly overexpressed in high glycerol consumption cultures. Moreover, *glpK* deletion has also been observed to be essential for glycerol utilization as the sole carbon source [69]. Then, deletion of this gene could result in a non-effective bioconversion process. As a result, this gene should not be taken into account for engineered *E. coli* strains using glycerol as the carbon source even when the GLYK reaction was repeatedly predicted to be knocked by OptKnock in ECC2 and iTA821 since two pathways for glycerol utilization in *E. coli* exist.

Based on OptKnock and random forest model predictions, four critical control points, glycolysis, pyruvate metabolism, the pentose phosphate pathway, and the TCA cycle, are associated with the overproduction of succinic acid. FUM and SUCDi appear to be the most significant keys in the TCA cycle for succinate overexpression. Results of this study suggest that they are mutually exclusive. Parallelly, the knockout of by-products such as acetate, formate, and lactate by deleting POX, ACKr, PTAr, PFL, and LDH_D were highly predicted to be knocked out. Those results are interesting since one of the bottlenecks for industrial production of bio-based products is the elimination of by-products, which could facilitate the recovery and purification process. This results and those obtained in the transcriptional responses suggest that deletion of the *pta* need to be, almost as mandatory, carried out since acetate production become a competitive pathway in glycerol metabolism for succinic acid production [68].

The pyruvate dehydrogenase complex is a critical connection point between glycolysis and the TCA cycle, both of which function during aerobic respiration through catalyzing the conversion of pyruvate to acetyl coenzyme A (acetyl-CoA) [56]. PDH deactivation results in PFL carrying the flux from pyruvate to acetyl-CoA [57]. Simple reaction knockouts show that PDH deletion results in a growth rate reduction of ~ 5%. Additionally, 5 reactions (FUM, GAPD, PGK, PGM, and TPI) were predicted to have the most significant reduction (8–10%) in the growth rate during glycerol utilization. Of those reactions, only FUM has a significant positive effect over the succinate production when this deletion is carried out alone. These results indicate that those mutants predicted by OptKnock where FUM and PDH are predicted are potential mutants to be tested in the lab because it have been observed that a low growth rate could negatively affect the profitability of industrial bio-based production products [2, 70].. However, in mutants in which both FUM and PDH were predicted (59.45%), TPI appeared in around 12.60% (Fig. 5). Then, the deletion of genes associated with TPI in addition to FUM and PDH reactions could negatively affect the growth rate. This because in the absence of TpiA, DHAP is converted to methylglyoxal, which, even at sub-millimolar concentrations, is a toxic compound [40]. DHAP is the results of the alternative pathway on glycerol metabolization consistent of an oxidation step by glycerol dehydrogenase (GldA). DHAP must be transformed into the general glycolytic pathway through isomerization by triosephosphate isomerase (TpiA) as glyceraldehyde-3-phosphate (GA3P). Therefore, deletion of *tpiA* could result in growth inhibition and cell death in the presence of glycerol as the only carbon source [69]. However, since FBA is not able to capture regulation, this situation could not be predicted by OptKnock.

Finally, computational models suggest that deletions of just 6–7 reaction knockouts are beneficial for industrial production since the growth rate does not decrease extremely. It is important to consider that a similar succinate production could be achieved if 6–8 reactions are knocked for all models. An assumption using optimization methods to predict cell capabilities is that the cell could quickly adjust the metabolism to maximize growth under certain conditions. This affirmation could be true for WT strains because FBA predicts an optimal condition. However, in metabolically engineered strains, the cell attempts to compensate the genetic changes carried out by the fewest changes in gene regulation until it achieves an optimal state that could be predicted using FBA [72]. Then, FBA in engineered strains predicts a long-term evolved state. Thus, an alternative to evaluate unevolved mutants is the MOMA method [24]. MOMA solves this problem by finding the solution that is most similar to the WT state (maximization of WT growth rate). Figure 7 shows a jump in the Euclidian distance between the WT and mutant strains when succinate production increases. This result could imply that after genetic manipulation microbial cell factories requires to be evolutionary engineered. ALE studies have shown to provide the cell with the ability to growth under selection pressure to go up from a sub-optimal state to the optimal growth rate predicted using *in silico* models [73]. Moreover, since OptKnock seeks to maximize the flux of a target chemical while maximizing the growth rate, our predictions could be beneficial for further ALE experiments because microbial cell factories have naturally evolved to maximize the growth rate. Thus, the succinic acid production rate would increase as biomass formation increases [55] by using ALE rounds after metabolically engineering cells [71].

Conclusions

By adopting tools from various disciplines, computational methods for systems metabolic engineering have been developed in order to understand the cell behavior and how level systems (RNA, proteins, metabolites, etc.) can interact inside the cell for industrial purposes. In the same way, *E. coli* has been extensively studied to become it a cell factory for the production of useful bio-based chemicals and materials through its native capabilities. However, there are some challenges that still need to be overcome.

This study proposes that computational tools have the potential to accelerate the optimization of cell factories by identifying metabolic engineering targets (genes/reactions) and not just by predicting mutants that may be biologically unviable. Therefore, systems metabolic engineering allows reduce time in rational strain design as well as to guide the selection of metabolic engineering targets based on the cell behavior under experimental conditions. At the same time, departing from traditional computational tools new methods such as machine learning could be proposed as an interesting alternative for the reduction of computational demand. However, these techniques are dependent on the level of completeness and accuracy of the metabolic model considered which could be improved by using omics data.

Methods

Strains and culture conditions

E. coli ATCC 8739 was obtained from the American Type Culture Collection (ATCC). A glycerol-based medium containing the following components (per liter) was used as the reference culture condition: 5 g of yeast extract, 2.5 g of NaCl, 5 mL of trace metal solution [0.55 g/L $CaCl_2$, 0.10 g/L $MnCl_2$ $4H_2O$, 0.17 g/L $ZnCl_2$, 0.043 g/L $CuCl_2$ $2H_2O$, 0.06 g/L $CoCl_2$ $6H_2O$, 0.06 g/L Na_2MoO_4 $2H_2O$, 0.06 g/L $Fe(NH_4)_2(SO4)_2$ $6H_2O$, 0.20 g/L $FeCl_3$ $6H_2O$], 5 mL of $MgSO_4$ (1 M), and 30 g of glycerol. For the optimized culture condition, the glycerol-based medium was supplemented with 1 g of NH_4Cl, 6 g of Na_2HPO_4, and 3 g of K_2HPO_4. For both conditions, a 50 mL culture was carried out in a 250 mL baffled conical Erlenmeyer flask and cultivated aerobically at 37 °C and 200 rpm.

Strains and culture conditions for adaptive laboratory evolution

To adapt *E. coli* ATCC 8739 on glycerol, *E. coli* was continuously sub-cultured in Luria-Bertani (LB) medium (5 g/L of NaCl, 10 g/L of tryptone, and 5 g/L of yeast extract) at 72 h per round until the 3rd round, where each round corresponds to one transfer to the same media. After every 3 rounds, the concentration of tryptone was decreased in 1 g/L until reaching 0 g/L of tryptone. Then, 10 rounds were carried out. A 50 mL culture was carried out in a 250 mL, non-baffled, conical Erlenmeyer flask and cultivated aerobically at 37 °C and 200 rpm. Then, 1 mL of the culture after each round was used as the inoculum for the next culture. Cells at the end of each ALE were kept at −80ºC and used for further evaluation of the growth and glycerol uptake, as well as transcriptomics analysis.

Differential expression analysis

RNA seq was carried out in triplicate for all conditions. For the evolved strain, the culture conditions for RNA-seq were the same as that for the optimized culture medium condition. To harvest the cells for total RNA purification, culture sample was first treated with RNAprotect Bacteria Reagent (Cat No./ID: 76506), and enzymatic lysis and Proteinase K digestion of the bacteria were carried out following the manufacturer's protocol. Then, the Qiagen RNeasy Mini kit (Cat No./ID: 74104), following the manufacturer's protocol, was used to obtain the total RNA for further analysis. Each sample was treated with DNase following the protocol in order to remove the DNA. The samples were sent to commercial RNA-seq services for further sample processing and sequencing (Genewiz, South Plainfield, NJ).

Clean, raw data was obtained by removing the reads containing adapters using trimmomatic. The sequence RefSeq: NC_CP010468 was employed for mapping. RNA reads were mapped using the software bowtie2, and featureCounts was employed for read counts. SARTools (Statistical Analysis of RNA-Seq data Tools) [74] was used for statistical RNA-Seq analysis. Differentially expressed genes (DEGs) were identified using the DESeq2 R

Package. The functional classification of the DEGs was performed using the gene ontology (GO) analysis by Blast2Go [39]. The data discussed in this publication has been deposited in NCBI's Gene Expression Omnibus [75] and are accessible through GEO Series accession number GSE140847.

Genome-scale metabolic network reconstruction

In order to obtain metabolic engineering targets to overproduce succinic acid from glycerol, the following two *E. coli* models were used: *EColiCore2* (ECC2) (data under peer review) and iTA1338 for *E. coli* ATCC 8739 (Supplementary File S1). Gene associations for both models were modified to ECOLC_RS number based on the sequence RefSeq: NC_CP010468 to facilitate the integration of transcriptomics data. Extensive manual curation was conducted, including (i) adding/eliminating transport reactions and extracellular metabolites and (ii) filling pathway gaps. Gap Find and Gap Filling, two optimization problems that search for root metabolite problems that are not connected in the network and that solve them, were used to fill gaps in iTA1338, including biomass reaction BIOMASS_Ec_iML1515_WT_75p37M (Supplementary File S1). All optimization problems were solved using the COBRA Toolbox v.3.0 [76].

Transcriptomics integration and metabolic engineering target prediction

Gene inactivity moderated by the metabolism and expression (GIMME) [77] method was used to integrate transcriptome data with the *E. coli* metabolic model. This method then minimized the usage of low-expression reactions while keeping the objective (e.g., biomass) above a certain value. Expressed genes were considered according to their expression level with Log_2 Fold Change (FC) $\geq |1|$. Next, according to the gene–protein reaction rules and the defined gene expression states, a specific activity state for each reaction was identified. Finally, a specific context model was obtained from the transcriptomic data. Metabolic engineering targets were obtained using OptKnock. The maximum uptake rate of glycerol was set to 13.3 mmol/gDW h^{-1}. The GIMME and OptKnock methods were conducted using COBRA Toolbox v.3.0 [76] in MATLAB 2017b and Gurobi 8.0.1.

Machine learning to determine potential metabolic engineering targets

Random forest models are supervised machine learning approaches, which have the advantage of giving a summary of the importance of each variable. This approach is based on a randomized variable selection process. An estimation of variable importance is provided by *IncNodePurity*, which measures the decrease in tree node purity that results from all splits of a given variable over all trees [78]. For interpretation purposes, this measure can be used to rank variables by the strength of their relation to the response variable [78]. A matrix of binary values was built from *m* mutant predicted and *n* reactions in the set of possible reactions to be knocked-out. In this matrix, one represents the presence of one specific reaction to be deleted in the mutant and zero absence in the combination of reactions to be deleted in the mutant. The matrix was partitioned into training and test sets; the training set was used to build a random forest model to predict succinic acid production, growth rate, or the growth rate Euclidean distance between the mutant and WT strains as response variables. For the training set, succinic acid production, growth rate variable response was initially predicted using FBA, and the growth rate Euclidean distance between the mutant and WT strains was predicted using MOMA. Next, the model performance was assessed using the testing set. Finally, we used the random forest to determine the importance of each target reaction over the three evaluated response variables.

Declarations

Funding

This research was results of the Gobernación del Cesar program of Science, Technology, and Innovation for higher education by PhD scholarships.

Author contributions

A.E.T.R: Conceptualization, Methodology, Software, Validation, Investigation, Visualization, Formal analysis, Writing-Original draft preparation. W.R: Investigation, D.M: Investigation. C.O: Investigation. R.C: Methodology, Software, Writing-Review & Editing. J.M.G.R: Supervision, Writing-Review & Editing. A.G.B: Conceptualization, Supervision, Writing-Review & Editing.

References

1. Vlysidis A, Binns M, Webb C, Theodoropoulos C. A techno-economic analysis of biodiesel biorefineries: Assessment of integrated designs for the co-production of fuels and chemicals. Energy [Internet]. 2011;36(8):4671–83. Available from: http://dx.doi.org/10.1016/j.energy.2011.04.046.
2. Chen X, Zhou L, Tian K, Kumar A, Singh S, Prior BA, et al. Metabolic engineering of Escherichia coli: A sustainable industrial platform for bio-based chemical production. Biotechnol Adv. 2013;31(8):1200–23.
3. Chen X, Xu G, Xu N, Zou W, Zhu P, Liu L, et al. Metabolic engineering of Torulopsis glabrata for malate production. Metab Eng [Internet]. 2013;19:10–6. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1096717613000505.
4. Woo HM, Park JB. Recent progress in development of synthetic biology platforms and metabolic engineering of Corynebacterium glutamicum. J Biotechnol [Internet]. 2014;180:43–51. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0168165614001060.
5. Kern A, Tilley E, Hunter IS, Legiša M, Glieder A. Engineering primary metabolic pathways of industrial micro-organisms. J Biotechnol [Internet]. 2007;129(1):6–29. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0168165606010182.
6. Förster AH, Gescher J. Metabolic engineering of Escherichia coli for production of mixed-acid fermentation end products. Front Bioeng Biotechnol. 2014;2(MAY):1–12.
7. Yin X, Li J, Shin H, dong, Du G, Liu L, Chen J. Metabolic engineering in the biotechnological production of organic acids in the tricarboxylic acid cycle of microorganisms: Advances and prospects. Biotechnol Adv [Internet]. 2015;33(6):830–41. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0734975015000713.
8. Zhang B, Skory CD, Yang ST. Metabolic engineering of Rhizopus oryzae: Effects of overexpressing pyc and pepc genes on fumaric acid biosynthesis from glucose. Metab Eng. 2012;14(5):512–20.
9. Shams Yazdani S, Gonzalez R. Engineering Escherichia coli for the efficient conversion of glycerol to ethanol and co-products. Metab Eng. 2008;10(6):340–51.
10. Zhu Q, Jackson EN. Metabolic engineering of Yarrowia lipolytica for industrial applications. Curr Opin Biotechnol [Internet]. 2015;36:65–72. Available from: http://dx.doi.org/10.1016/j.copbio.2015.08.010.
11. Buschke N, Schäfer R, Becker J, Wittmann C. Metabolic engineering of industrial platform microorganisms for biorefinery applications - Optimization of substrate spectrum and process robustness by rational and evolutive

strategies. Bioresour Technol [Internet]. 2013;135:544–54. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0960852412017294.

12. Rangel AET, Gómez Ramírez JM, González Barrios AF. From industrial by-products to value-added compounds: the design of efficient microbial cell factories by coupling systems metabolic engineering and bioprocesses. Biofuels, Bioprod Biorefining. 2020;1–11.
13. O'Brien EJ, Monk JM, Palsson BO. Using genome-scale models to predict biological capabilities. Cell [Internet]. 2015 [cited 2017 Mar 6];161(5):971–87. Available from: http://dx.doi.org/10.1016/j.cell.2015.05.019.
14. Haggart CR, Bartell JA, Saucerman JJ, Papin JA. Whole-genome metabolic network reconstruction and constraint-based modeling. Methods Enzymol [Internet]. 2011 [cited 2019 Jan 21];500:411–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21943909.
15. Carlson R, Fell D, Srienc F. Metabolic pathway analysis of a recombinant yeast for rational strain development. Biotechnol Bioeng. 2002;79(2):121–34.
16. Portela C, Villas-Bôas S, Rocha I, Ferreira EC. Genome scale metabolic network reconstruction of the pathogen Enterococcus faecalis. In: IFAC Proceedings Volumes (IFAC-PapersOnline) [Internet]. Elsevier; 2013 [cited 2019 Jan 21]. p. 131–6. Available from: https://www.sciencedirect.com/science/article/pii/S1474667016313830.
17. Pharkya P, Maranas CD. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng. 2006;8(1):1–13.
18. Kim J. Development and Applications of Integrated Metabolic and Transcriptional Regulatory Network Models. UNIVERSITY OF WISCONSIN–MADISON; 2012.
19. Ruckerbauer DE, Jungreuthmayer C, Zanghellini J. Predicting genetic engineering targets with Elementary Flux Mode Analysis: A review of four current methods. N Biotechnol. 2015;32(6):534–46.
20. Pharkya P, Burgard AP, Maranas CD. OptStrain: A computational framework for redesign of microbial production systems. Genome Res [Internet]. 2004 Nov [cited 2016 Oct 5];14(11):2367–76. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15520298.
21. Ranganathan S, Suthers PF, Maranas CD. OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput Biol [Internet]. 2010 Apr [cited 2016 May 18];6(4):e1000744. Available from: http://journals.plos.org/ploscompbiol/article?id=10.1371%2Fjournal.pcbi.1000744.
22. Burgard AP, Pharkya P, Maranas CD. OptKnock:. A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization. Biotechnol Bioeng [Internet]. 2003 Dec 20 [cited 2015 Mar 11];84(6):647–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14595777.
23. Pharkya P, Burgard AP, Maranas CD. Exploring the Overproduction of Amino Acids Using the Bilevel Optimization Framework OptKnock. Biotechnol Bioeng [Internet]. 2003 Dec 30 [cited 2016 May 22];84(7):887–99. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14708128.
24. Segre D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci [Internet]. 2002 Nov 12 [cited 2018 Jun 11];99(23):15112–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12415116.
25. Nordlander B, Krantz M, Hohmann S. Hog1-mediated metabolic adjustments following hyperosmotic shock in the yeast. Current [Internet]. 2008;20(August 2007):51–79. Available from: http://dx.doi.org/10.1007/4735_2007_0256.

26. Blazier AS, Papin JA. Integration of expression data in genome-scale metabolic network reconstructions. Front Physiol. 2012;3 AUG(August).
27. Feist AM, Zielinski DC, Orth JD, Schellenberger J, Herrgard MJ, Palsson BO. Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. Metab Eng. 2010;12(3):173–86.
28. Machado D, Herrgård M. Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. PLoS Comput Biol [Internet]. 2014 [cited 2017 Mar 21];10(4). Available from: http://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1003580&type=printable.
29. Conrad TM, Lewis NE, Palsson BO. Microbial laboratory evolution in the era of genome-scale science. Mol Syst Biol. 2011 Jul 5;7(1):509–509.
30. Lee DH, Palsson BO. Adaptive evolution of escherichia coli K-12 MG1655 during growth on a nonnative carbon source, L-l,2-propanediol. Appl Environ Microbiol [Internet]. 2010 [cited 2017 Mar 9];76(13):4158–68. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2897412/pdf/0373-10.pdf.
31. Wang Y, Manow R, Finan C, Wang J, Garza E, Zhou S. Adaptive evolution of nontransgenic Escherichia coli KC01 for improved ethanol tolerance and homoethanol fermentation from xylose. J Ind Microbiol Biotechnol [Internet]. 2011 [cited 2017 Mar 9];38(9):1371–7. Available from: http://download.springer.com.ezproxy.uniandes.edu.co:8080/static/pdf/245/art%253A10.1007%252Fs10295-010-0920-5.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs10295-010-0920-5&token2=exp=1489072164~acl=%2Fstatic%2Fpdf%2F245%2Fart%25253.
32. http://genome.cshlp.org/cgi/content/abstract/15/10/1365%5Cnpapers://9c79d1e6-ac2d-4d83-b796-6cea7716b2a4/Paper/p403.
33. Fong SS, Marciniak JY. Description and Interpretation of Adaptive Evolution of Escherichia coli K-12 MG1655 by Using a Genome-Scale In Silico Metabolic Model. Society [Internet]. 2003 [cited 2017 Mar 9];185(21):6400–8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC219384/pdf/0253.pdf.
34. Bao H, Liu R, Liang L, Jiang Y, Jiang M, Ma J, et al. Succinic acid production from hemicellulose hydrolysate by an Escherichia coli mutant obtained by atmospheric and room temperature plasma and adaptive evolution. Enzyme Microb Technol [Internet]. 2014 Nov [cited 2017 Mar 9];66:10–5. Available from: http://ac.els-cdn.com.ezproxy.uniandes.edu.co:8080/S0141022914000908/1-s2.0-S0141022914000908-main.pdf?_tid=55ae8ed0-04f5-11e7-92d5-00000aab0f6b&acdnat=1489083921_347e3210792409f284efbe8c39db3739.
35. Zhang J, Wu C, Du G, Chen J. Enhanced acid tolerance in Lactobacillus casei by adaptive evolution and compared stress response during acid stress. Biotechnol Bioprocess Eng [Internet]. 2012 [cited 2017 Mar 9];17(2):283–9. Available from: http://download.springer.com.ezproxy.uniandes.edu.co:8080/static/pdf/799/art%253A10.1007%252Fs12257-011-0346-6.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Farticle%2F10.1007%2Fs12257-011-0346-6&token2=exp=1489072050~acl=%2Fstatic%2Fpdf%2F799%2Fart%25253.
36. Lee JW, Na D, Park JM, Lee J, Choi S, Lee SY. Systems metabolic engineering of microorganisms for natural and non-natural chemicals. Nat Chem Biol. 2012;8(6):536–46.
37. Furusawa C, Horinouchi T, Hirasawa T, Shimizu H. Systems metabolic engineering: The creation of microbial cell factories by rational metabolic design and evolution. Adv Biochem Eng Biotechnol. 2013;131:1–23.
38. Makino K, Shinagawa H, Amemura M, Kawamoto T, Yamada M, Nakata A. Signal transduction in the phosphate regulon of Escherichia coli involves phosphotransfer between PhoR and PhoB proteins. J Mol Biol. 1989;210(3):551–9.

39. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008;36(10):3420–35.
40. Booth IR. Glycerol and Methylglyoxal Metabolism. EcoSal Plus. 2014;1(2):1–8.
41. Kim YM, Cho HS, Jung GY, Park JM. Engineering the pentose phosphate pathway to improve hydrogen yield in recombinant Escherichia coli. Biotechnol Bioeng. 2011;108(12):2941–6.
42. Jones HM, Gunsalus RP. Regulation of Escherichia coli fumarate reductase (frdABCD) operon expression by respiratory electron acceptors and the fnr gene product. J Bacteriol. 1987;169(7):3340–9.
43. Cecchini G, Schröder I, Gunsalus RP, Maklashina E. Succinate dehydrogenase and fumarate reductase from Escherichia coli. Biochim Biophys Acta - Bioenerg. 2002;1553(1–2):140–57.
44. Chen YP, Lin HH, Yang CD, Huang SH, Tseng CP. Regulatory role of cAMP receptor protein over Escherichia coli fumarase genes. J Microbiol. 2012;50(3):426–33.
45. Chagneau C, Heyde M, Alonso S, Portalier R, Laloi P. External-pH-dependent expression of the maltose regulon and ompF gene in Escherichia coli is affected by the level of glycerol kinase, encoded by glpK. J Bacteriol. 2001;183(19):5675–83.
46. Blankschien MD, Clomburg JM, Gonzalez R. Metabolic engineering of Escherichia coli for the production of succinate from glycerol. Metab Eng [Internet]. 2010;12(5):409–19. Available from: http://dx.doi.org/10.1016/j.ymben.2010.06.002.
47. Lin H, Bennett GN, San K-Y. Metabolic engineering of aerobic succinate production systems in Escherichia coli to improve process productivity and achieve the maximum theoretical succinate yield. Metab Eng [Internet]. 2005;7(2):116–27. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1096717604000783.
48. Ataman M, Hernandez Gardiol DF, Fengos G, Hatzimanikatis V. redGEM: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models. Senger RS, editor. PLoS Comput Biol [Internet]. 2017 Jul 20 [cited 2019 Sep 30];13(7):e1005444. Available from: https://dx.plos.org/10.1371/journal.pcbi.1005444.
49. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, et al. Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. Proc Natl Acad Sci U S A [Internet]. 2013 [cited 2019 Sep 17];110(50):20338–43. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3864276/.
50. Monk JM, Koza A, Campodonico MA, Machado D, Seoane JM, Palsson BO, et al. Multi-omics Quantification of Species Variation of Escherichia coli Links Molecular Features with Strain Phenotypes. Cell Syst [Internet]. 2016 Sep 28 [cited 2018 Jun 7];3(3):238–251.e12. Available from: https://www.sciencedirect.com/science/article/pii/S2405471216302903?via%3Dihub.
51. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, et al. iML1515, a knowledgebase that computes Escherichia coli traits. Nat Biotechnol [Internet]. 2017 Oct 11 [cited 2017 Oct 25];35(10):904–8. Available from: http://www.nature.com/doifinder/10.1038/nbt.3956.
52. Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, et al. Experimental and computational assessment of conditionally essential genes in Escherichia coli. J Bacteriol [Internet]. 2006 Dec [cited 2019 Jan 28];188(23):8259–71. Available from: papers3://publication/doi/10.1128/JB.00740-06.
53. López-Garzón CS, Straathof AJJ. Recovery of carboxylic acids produced by fermentation. Biotechnol Adv. 2014;32(5):873–904.

54. Kurzrock T, Weuster-Botz D. Recovery of succinic acid from fermentation broth. Biotechnol Lett. 2010;32(3):331–9.
55. Shabestary K, Hudson EP. Computational metabolic engineering strategies for growth-coupled biofuel production by Synechocystis. Metab Eng Commun [Internet]. 2016 Dec [cited 2019 Sep 24];3:216–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29468126.
56. Schutte KM, Fisher DJ, Burdick MD, Mehrad B, Mathers AJ, Mann BJ, et al. Escherichia coli pyruvate dehydrogenase complex is an important component of CXCL10-mediated antimicrobial activity. Infect Immun [Internet]. 2015 [cited 2019 Sep 24];84(1):320–8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4694015/.
57. Khodayari A, Chowdhury A, Maranas CD. Succinate Overproduction: A Case Study of Computational Strain Design Using a Comprehensive Escherichia coli Kinetic Model. Front Bioeng Biotechnol [Internet]. 2015 Jan 5 [cited 2016 May 18];2:76. Available from: http://journal.frontiersin.org/article/10.3389/fbioe.2014.00076/abstract.
58. Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, et al. Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale. Nat Genet. 2006;38(12):1406–12.
59. Wong KK, Kwan HS. Transcription of glpT of Escherichia coli K12 is regulated by anaerobiosis and fnr. FEMS Microbiol Lett [Internet]. 1992 Jul 1 [cited 2019 Sep 16];94(1–2):15–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1521763.
60. Lemieux MJ, Huang Y, Wang DN. Crystal structure and mechanism of GlpT, the glycerol-3-phosphate transporter from E. coli. J Electron Microsc (Tokyo). 2005;54(SUPPL. 1):43–6.
61. Gao R, Stock AM. Evolutionary Tuning of Protein Expression Levels of a Positively Autoregulated Two-Component System. PLoS Genet. 2013;9(10).
62. Baek JH, Lee SY. Transcriptome analysis of phosphate starvation response in Escherichia coli. J Microbiol Biotechnol. 2007;17(2):244–52.
63. Unden G, Bongaerts J. Alternative respiratory pathways of Escherichia coli: Energetics and transcriptional regulation in response to electron acceptors. Biochim Biophys Acta - Bioenerg. 1997;1320(3):217–34.
64. Murarka A, Dharmadi Y, Yazdani SS, Gonzalez R. Fermentative utilization of glycerol by Escherichia coli and its implications for the production of fuels and chemicals. Appl Environ Microbiol [Internet]. 2008 Feb [cited 2016 Nov 17];74(4):1124–35. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18156341.
65. Guest JR. Partial replacement of succinate dehydrogenase function by phage- and plasmid-specified fumarate reductase in Escherichia coli. J Gen Microbiol. 1981;122(2):171–9.
66. Masuda N, Church GM. Regulatory network of acid resistance genes in Escherichia coli. Mol Microbiol. 2003;48(3):699–712.
67. Kern R, Malki A, Abdallah J, Tagourti J, Richarme G. Escherichia coli HdeB is an acid stress chaperone. J Bacteriol. 2007;189(2):603–10.
68. Zhang X, Shanmugam KT, Ingram LO. Fermentation of glycerol to succinate by metabolically engineered strains of escherichia coli. Appl Environ Microbiol [Internet]. 2010 Apr 15 [cited 2016 Feb 8];76(8):2397–401. Available from: http://aem.asm.org/content/76/8/2397.full.
69. Velur Selvamani RS, Telaar M, Friehs K, Flaschel E. Antibiotic-free segregational plasmid stabilization in Escherichia coli owing to the knockout of triosephosphate isomerase (tpiA). Microb Cell Fact [Internet]. 2014

Apr 21 [cited 2019 Sep 24];13(1):58. Available from: http://microbialcellfactories.biomedcentral.com/articles/10.1186/1475-2859-13-58.

70. Tafur Rangel AE, Camelo Valera LC, Gómez Ramírez JM, González Barrios AF. Effects of metabolic engineering on downstream processing operational cost and energy consumption: the case of Escherichia coli's glycerol conversion to succinic acid. J Chem Technol Biotechnol. 2018;93(7):2011–20.
71. Graf M, Haas T, Müller F, Buchmann A, Harm-Bekbenbetova J, Freund A, et al. Continuous Adaptive Evolution of a Fast-Growing Corynebacterium glutamicum Strain Independent of Protocatechuate. Front Microbiol [Internet]. 2019 [cited 2019 Sep 24];10:1648. Available from: http://www.ncbi.nlm.nih.gov/pubmed/31447790.
72. Senger R, Yen J, Tanniche I, Fisher A, Gillaspy G, Bevan D. Designing metabolic engineering strategies with genome-scale metabolic flux modeling. Adv Genomics Genet. 2015;93.
73. Ibarra RU, Edwards JS, Palsson BO. Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature. 2002;420(6912):186–9.
74. Varet H, Brillet-Guéguen L, Coppée JY, Dillies MA. SARTools:. A DESeq2- and edgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. Mills K, editor. PLoS One [Internet]. 2016 Jun 9 [cited 2018 Jul 3];11(6):e0157022. Available from: http://dx.plos.org/10.1371/journal.pone.0157022.
75. Edgar R. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30(1):207–10.
76. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. Nat Protoc [Internet]. 2019 Mar 20 [cited 2019 Sep 17];14(3):639–702. Available from: http://www.nature.com/articles/s41596-018-0098-2.
77. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol. 2008;4(5).
78. Li J, Poursat MA, Drubay D, Motz A, Saci Z, Morillon A, et al. A Dual Model for Prioritizing Cancer Mutations in the Non-coding Genome Based on Germline and Somatic Events. PLoS Comput Biol. 2015;11(11):1–17.
79. King ZA, Lu J, Dräger A, Miller P, Federowicz S, Lerman JA, et al. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res [Internet]. 2016 Jan 4 [cited 2017 Apr 18];44(D1):D515–22. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkv1049.

Figures

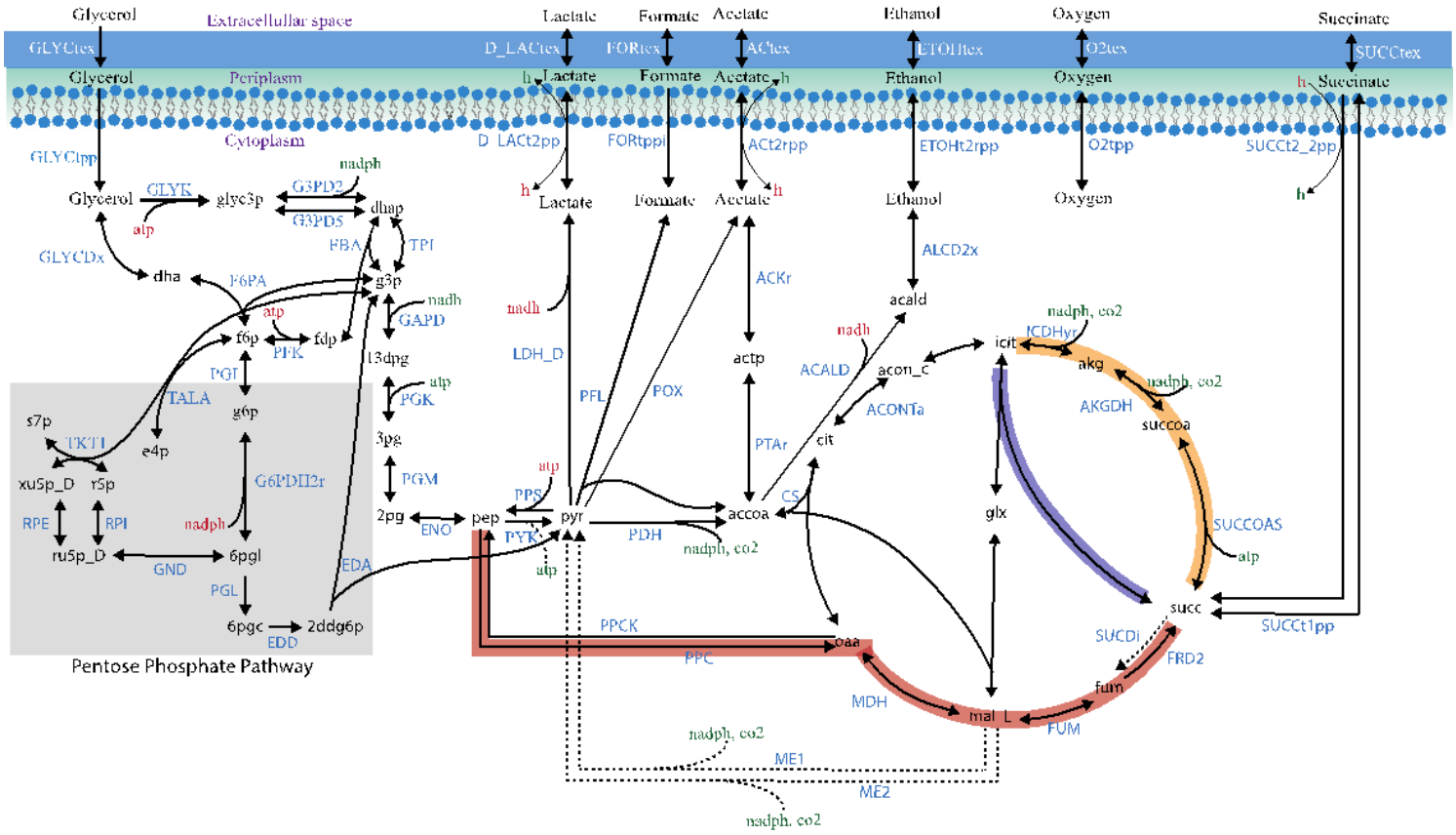

Figure 1

Succinic acid pathways from glycerol in Escherichia coli. The three pathways for succinic acid production are indicated by the thick red (the PEP–pyruvate–oxaloacetate node -the reductive TCA branch-), yellow (the oxidative TCA branch), and blue (the glyoxylate shunt) arrows. Relevant biochemical reactions are represented based on the id BIGG database names [79].

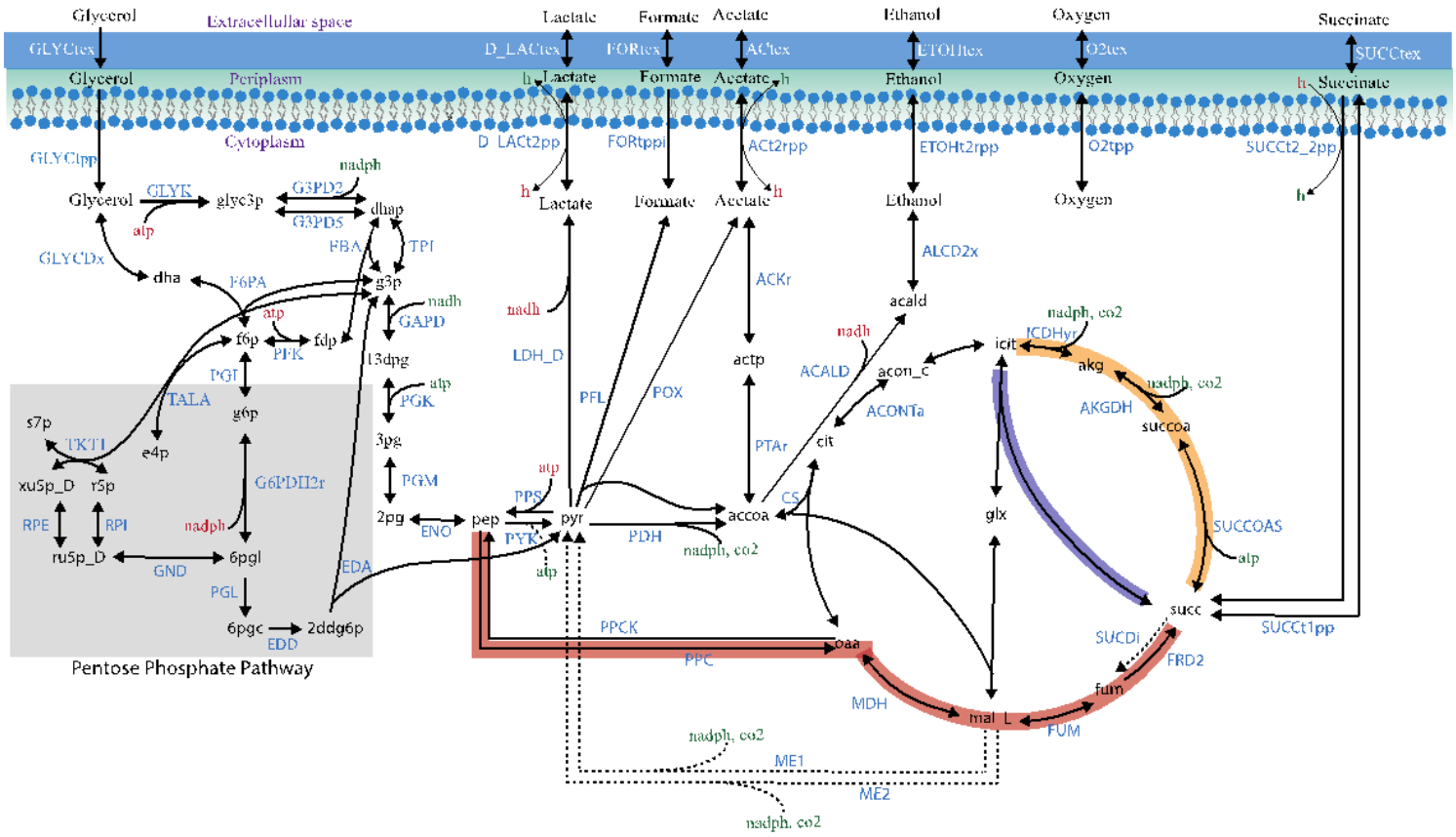

Figure 1

Succinic acid pathways from glycerol in Escherichia coli. The three pathways for succinic acid production are indicated by the thick red (the PEP–pyruvate–oxaloacetate node -the reductive TCA branch-), yellow (the oxidative TCA branch), and blue (the glyoxylate shunt) arrows. Relevant biochemical reactions are represented based on the id BIGG database names [79].

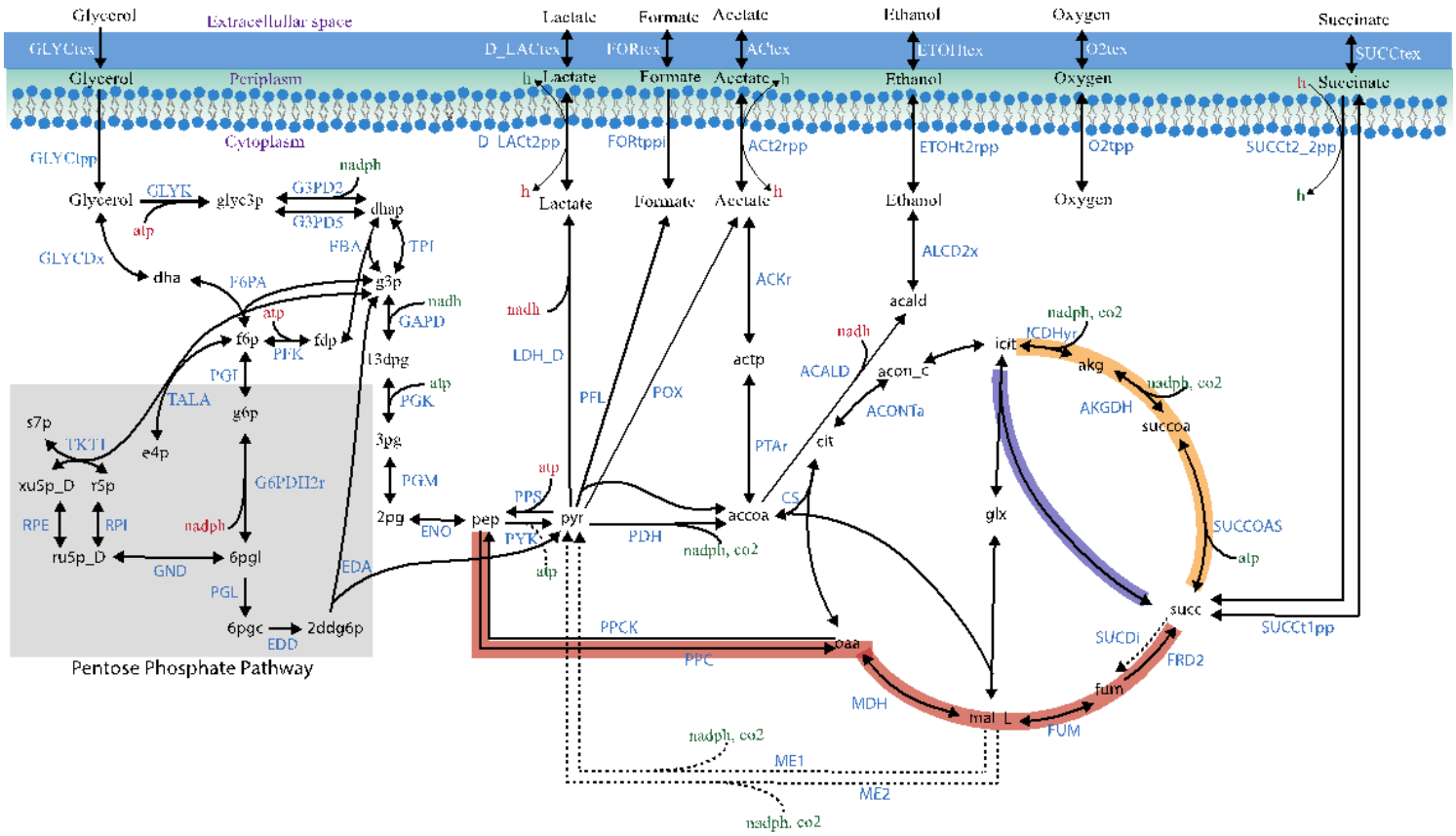

Figure 1

Succinic acid pathways from glycerol in Escherichia coli. The three pathways for succinic acid production are indicated by the thick red (the PEP–pyruvate–oxaloacetate node -the reductive TCA branch-), yellow (the oxidative TCA branch), and blue (the glyoxylate shunt) arrows. Relevant biochemical reactions are represented based on the id BIGG database names [79].

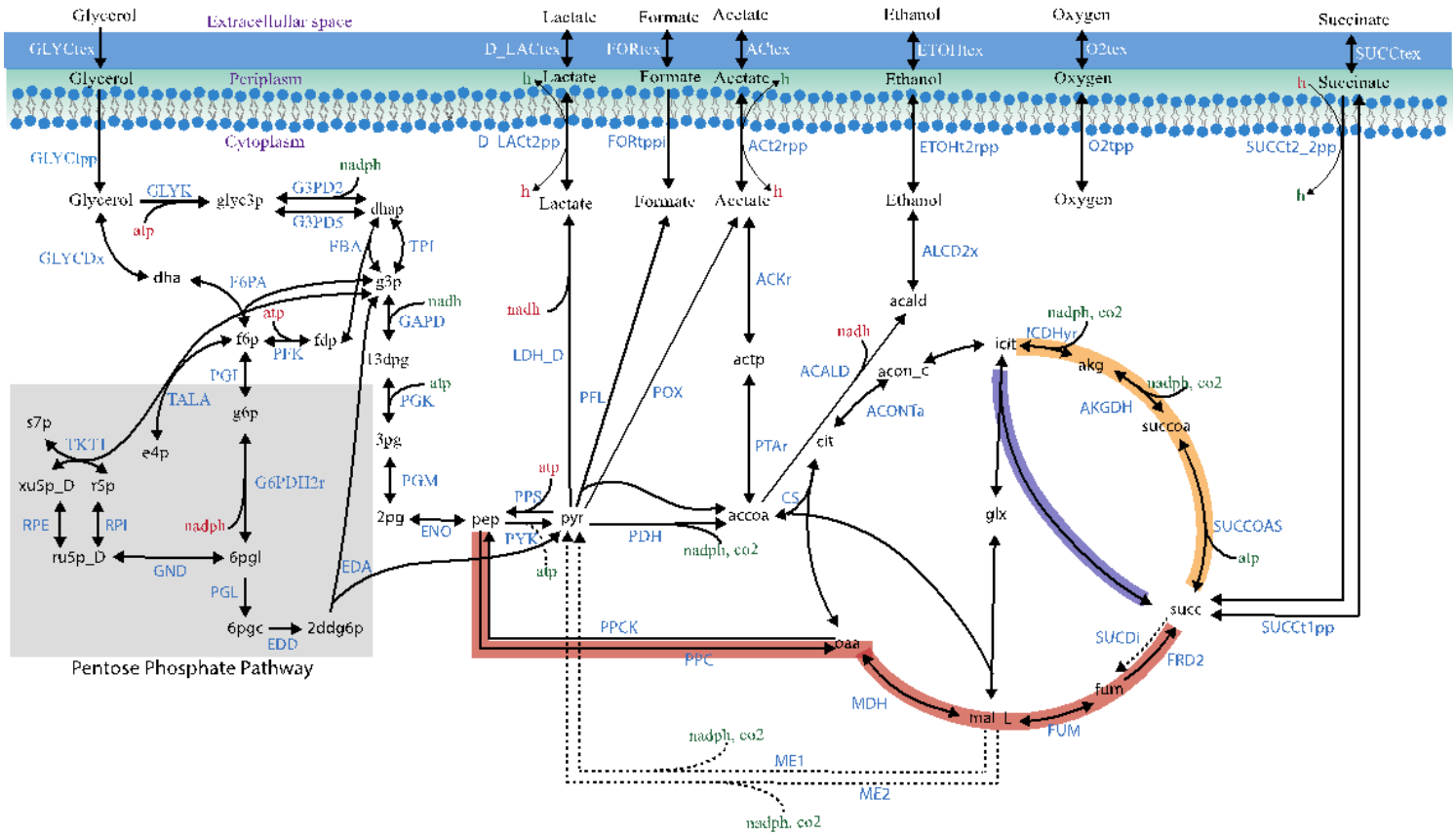

Figure 1

Succinic acid pathways from glycerol in Escherichia coli. The three pathways for succinic acid production are indicated by the thick red (the PEP–pyruvate–oxaloacetate node -the reductive TCA branch-), yellow (the oxidative TCA branch), and blue (the glyoxylate shunt) arrows. Relevant biochemical reactions are represented based on the id BIGG database names [79].

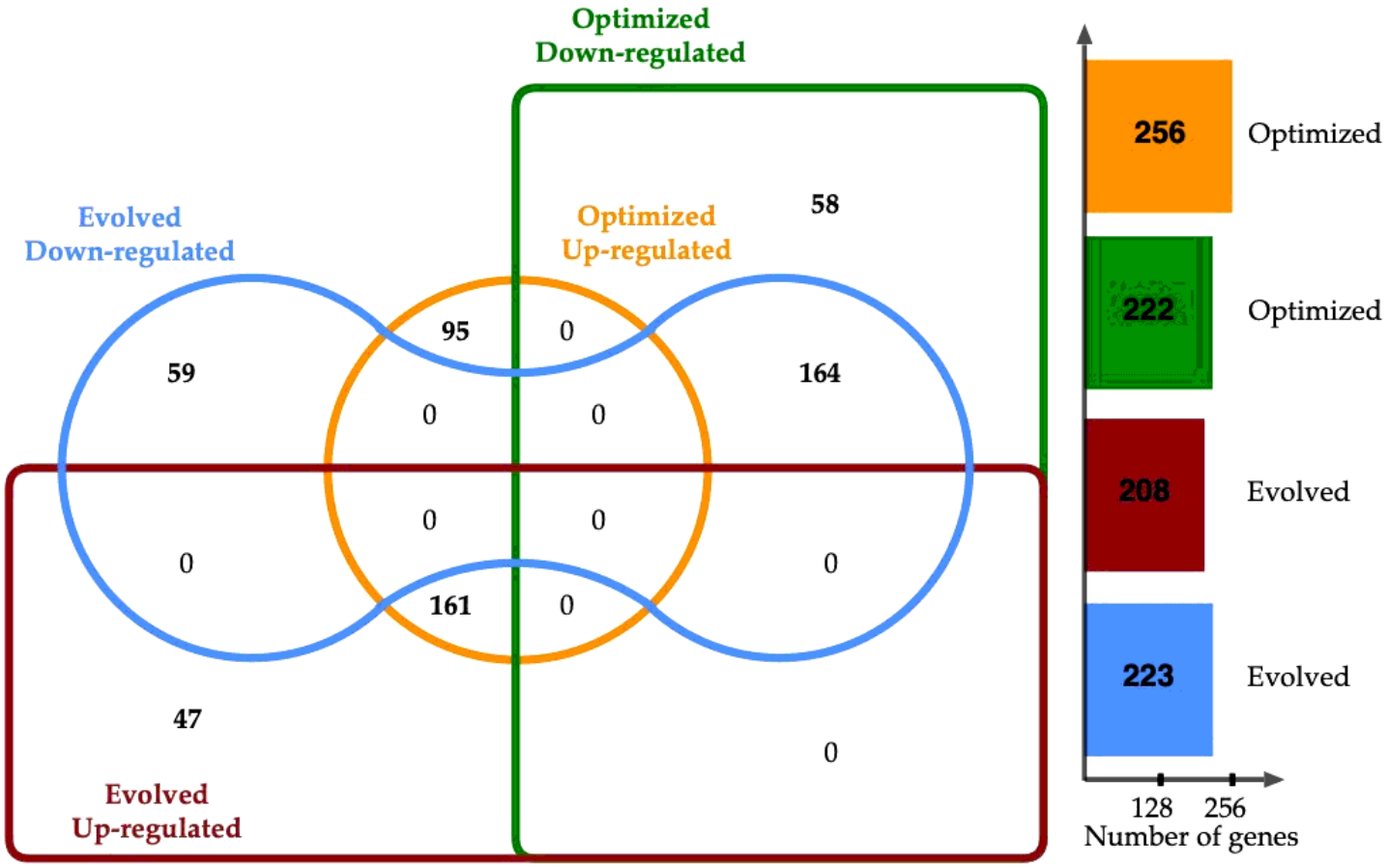

Figure 2

E. coli ATCC 8739 differentially expressed genes in the optimized medium vs glycerol-based medium and the evolved strain vs glycerol-based medium.

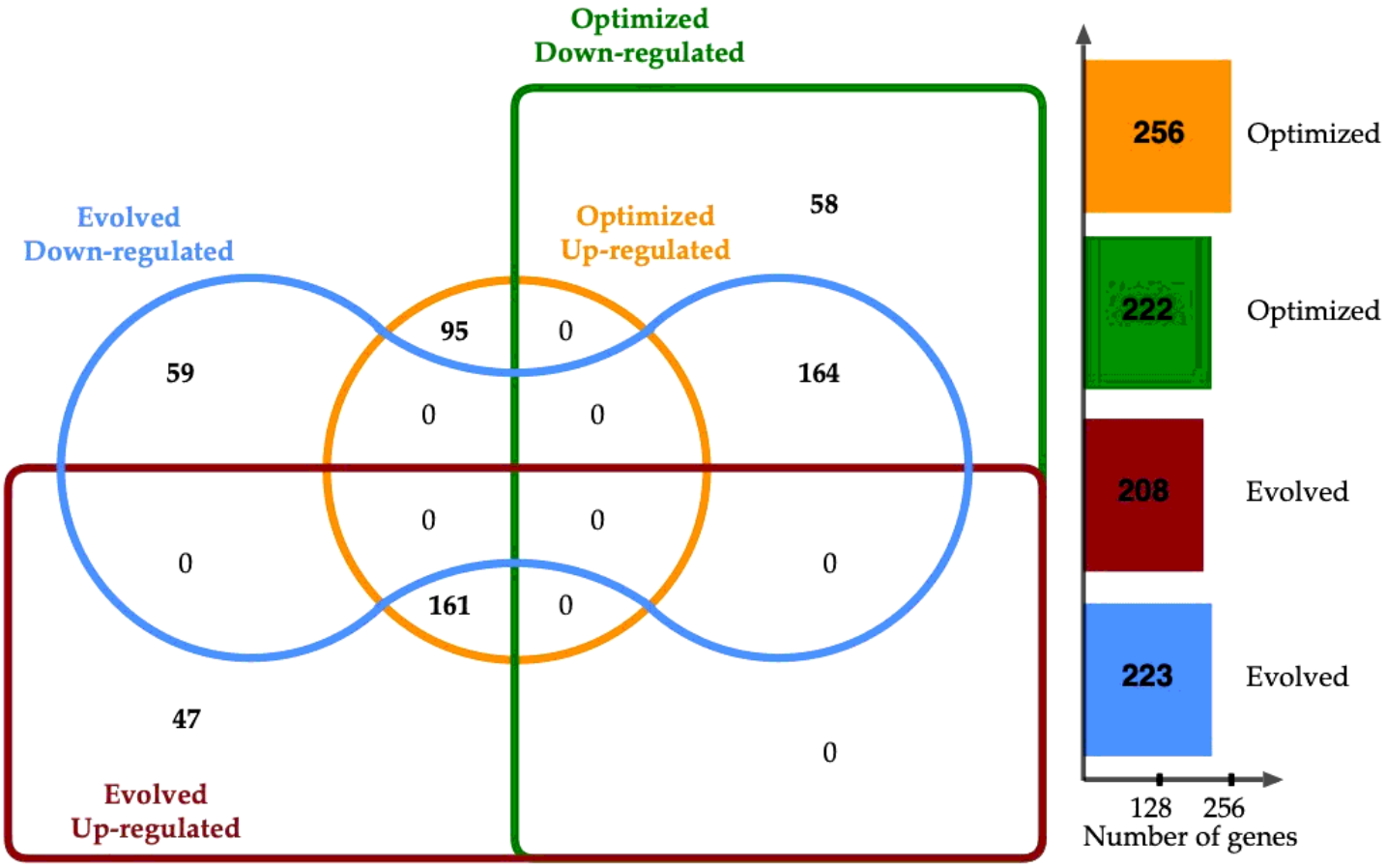

Figure 2

E. coli ATCC 8739 differentially expressed genes in the optimized medium vs glycerol-based medium and the evolved strain vs glycerol-based medium.

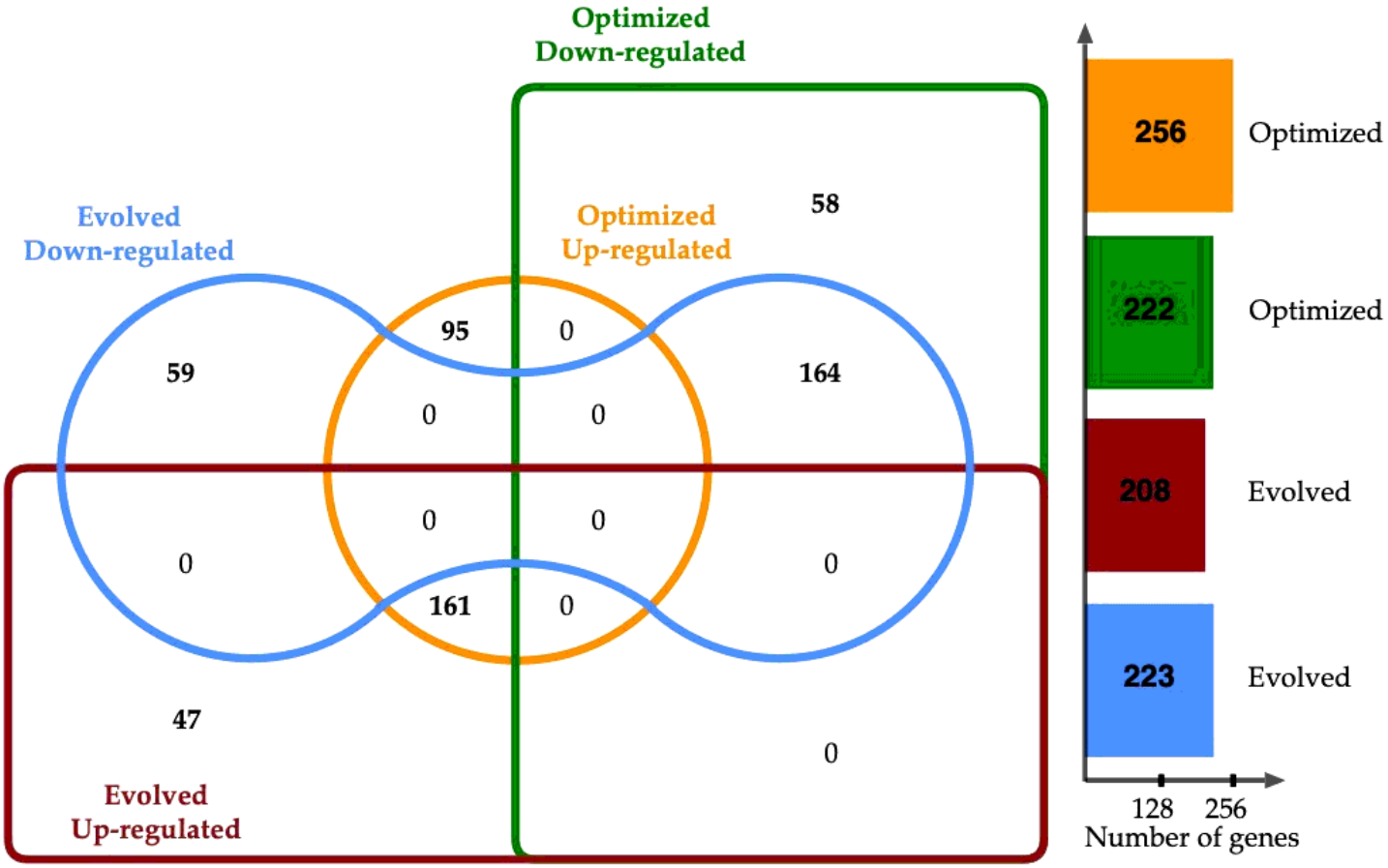

Figure 2

E. coli ATCC 8739 differentially expressed genes in the optimized medium vs glycerol-based medium and the evolved strain vs glycerol-based medium.

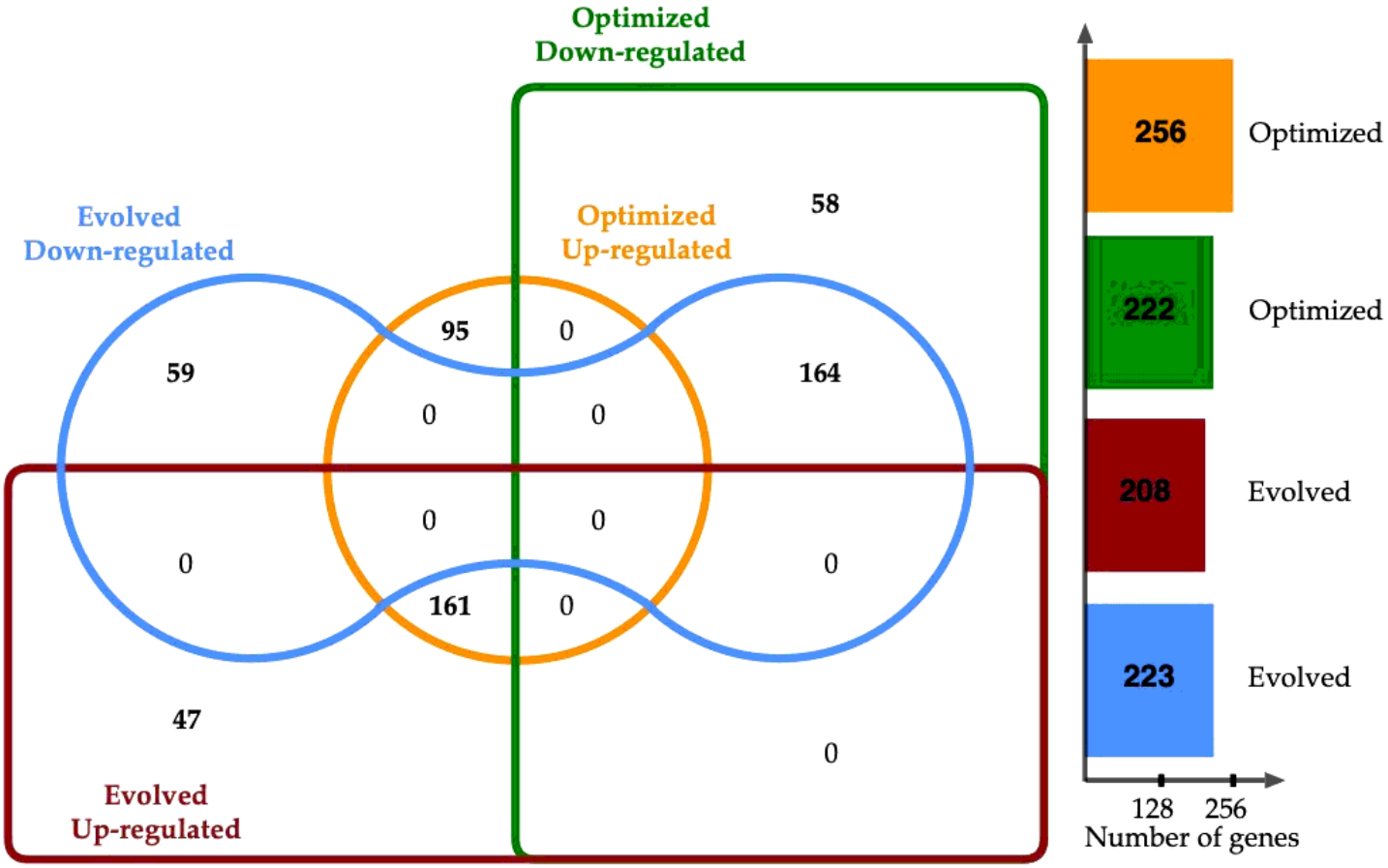

Figure 2

E. coli ATCC 8739 differentially expressed genes in the optimized medium vs glycerol-based medium and the evolved strain vs glycerol-based medium.

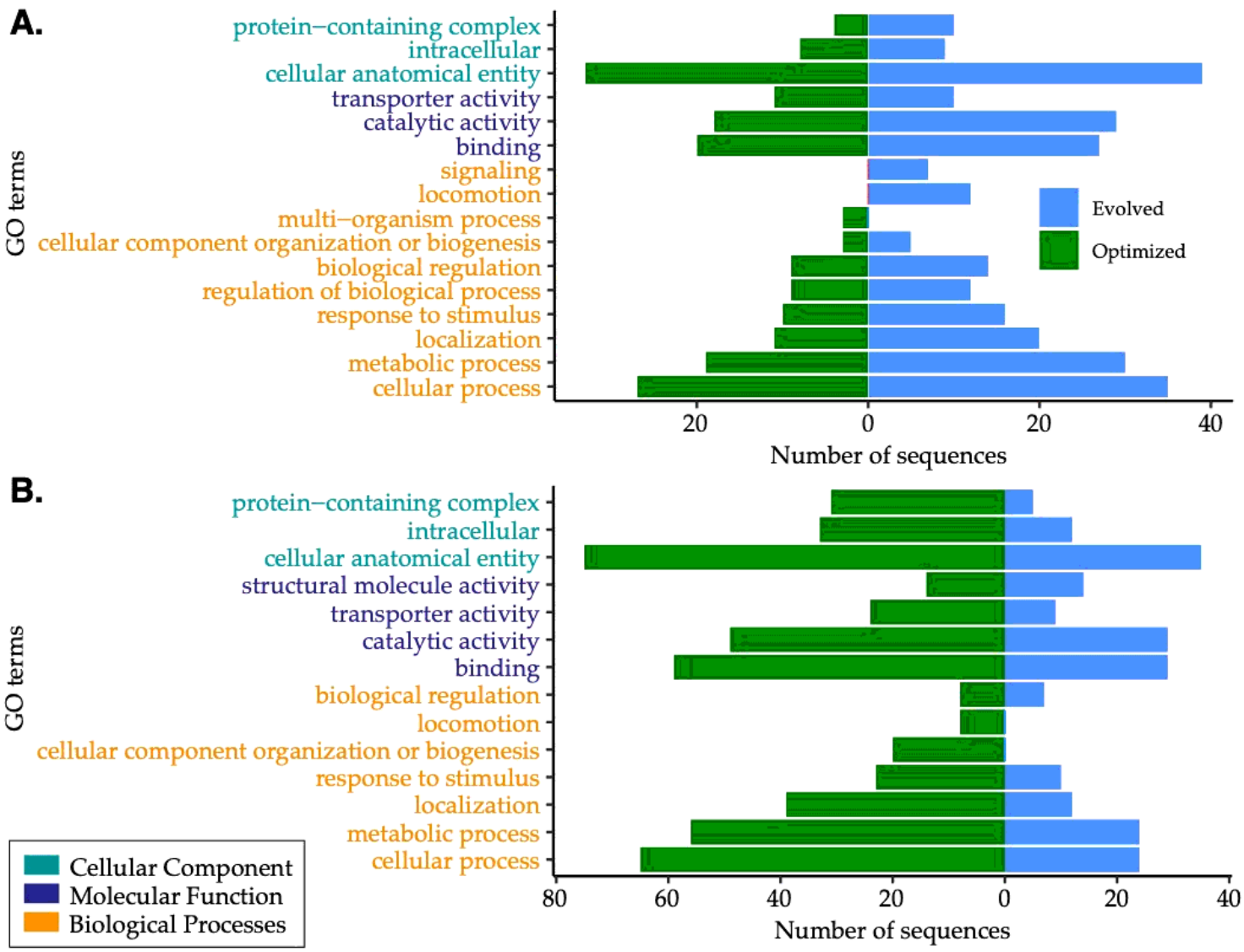

Figure 3

Gene ontology (GO) analysis of unique differentially expressed genes. A. Down regulated genes B. Up-regulated genes.

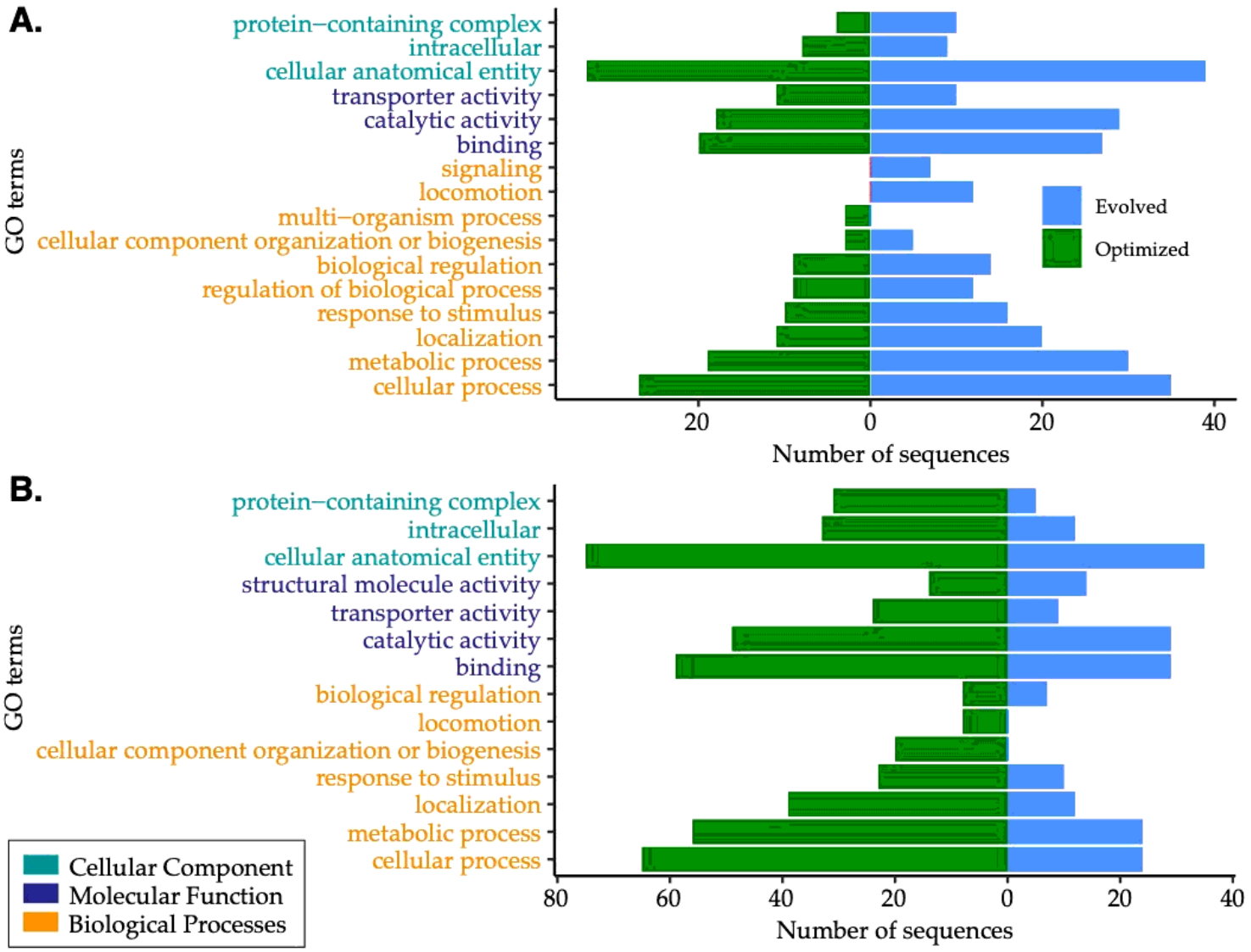

Figure 3

Gene ontology (GO) analysis of unique differentially expressed genes. A. Down regulated genes B. Up-regulated genes.

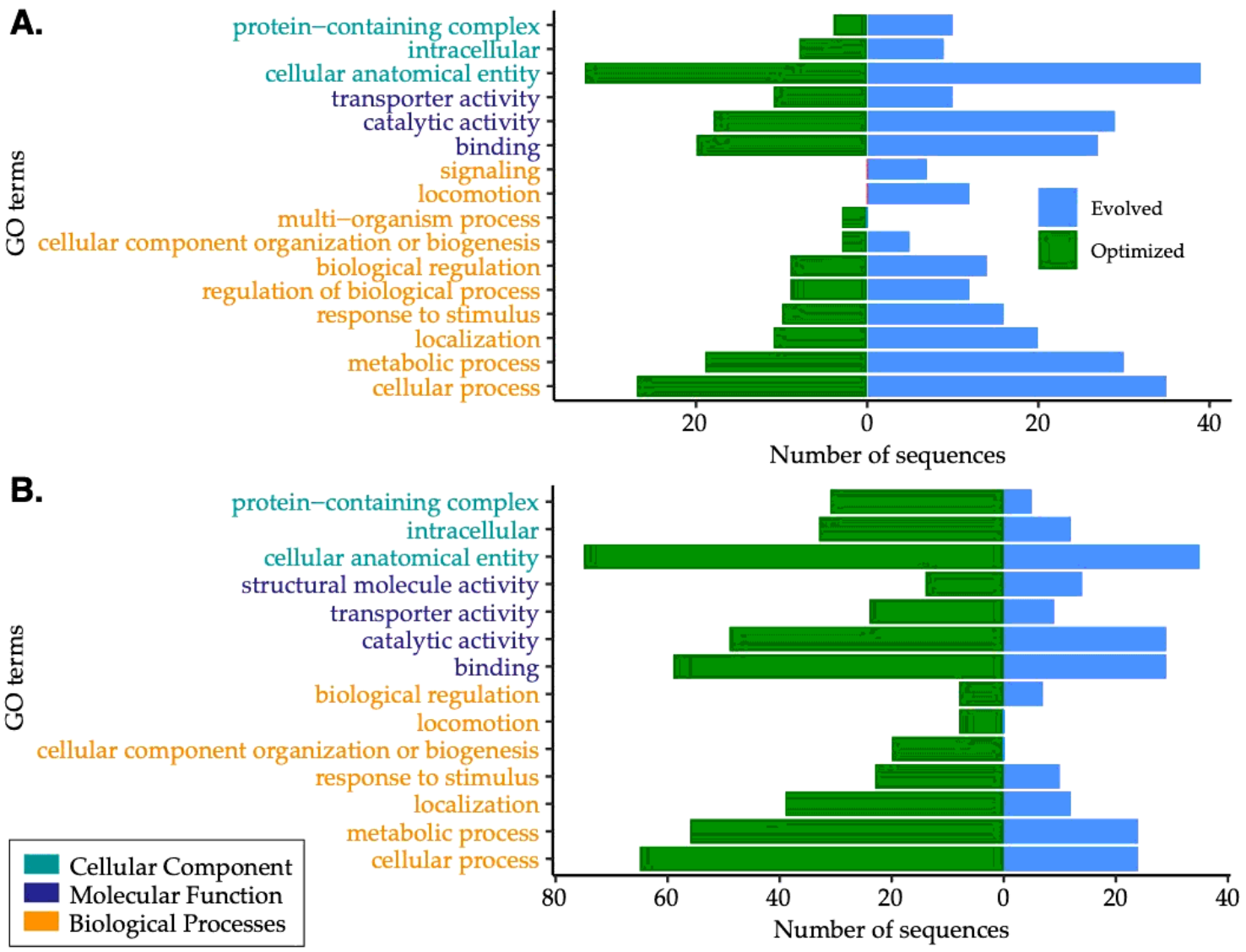

Figure 3

Gene ontology (GO) analysis of unique differentially expressed genes. A. Down regulated genes B. Up-regulated genes.

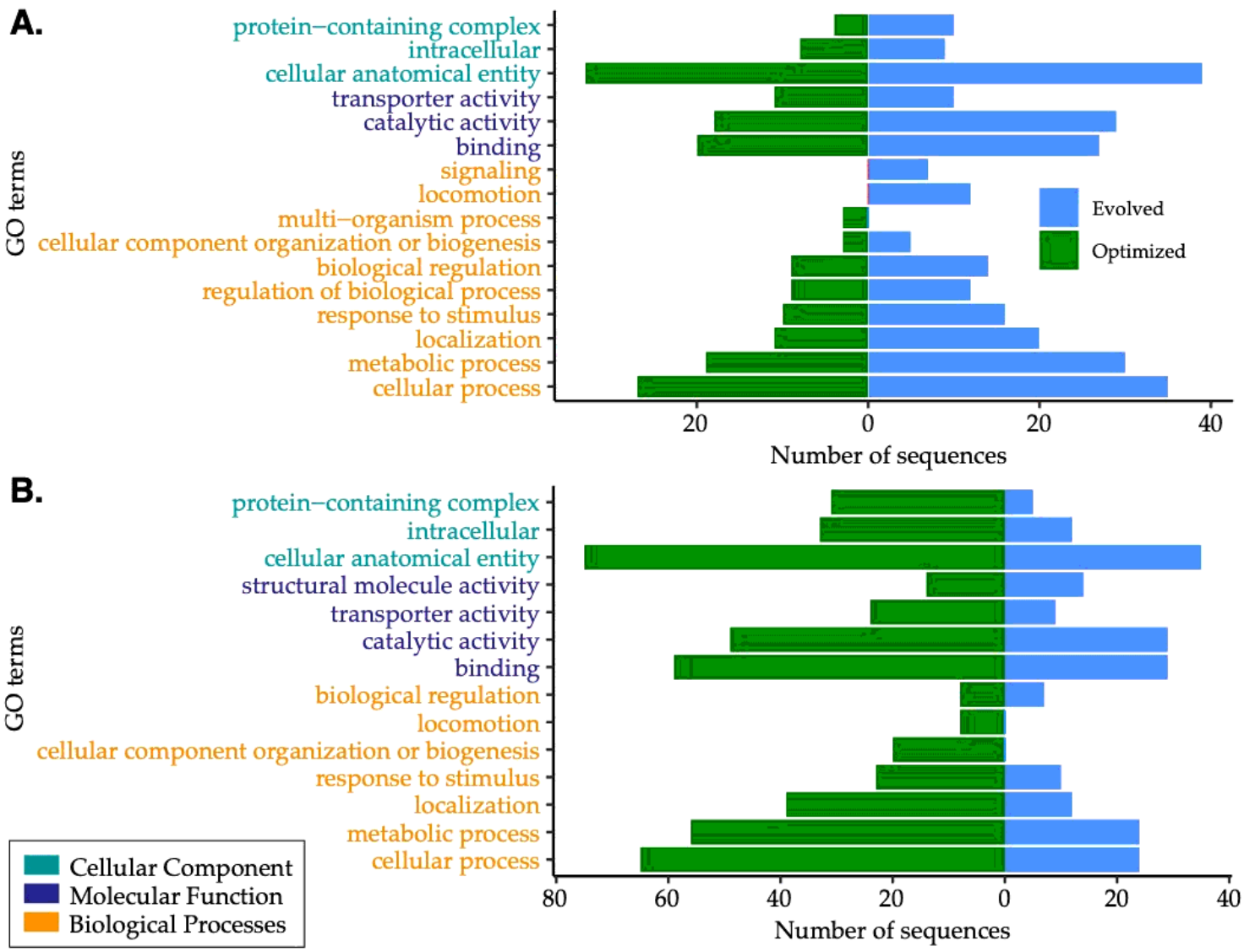

Figure 3

Gene ontology (GO) analysis of unique differentially expressed genes. A. Down regulated genes B. Up-regulated genes.

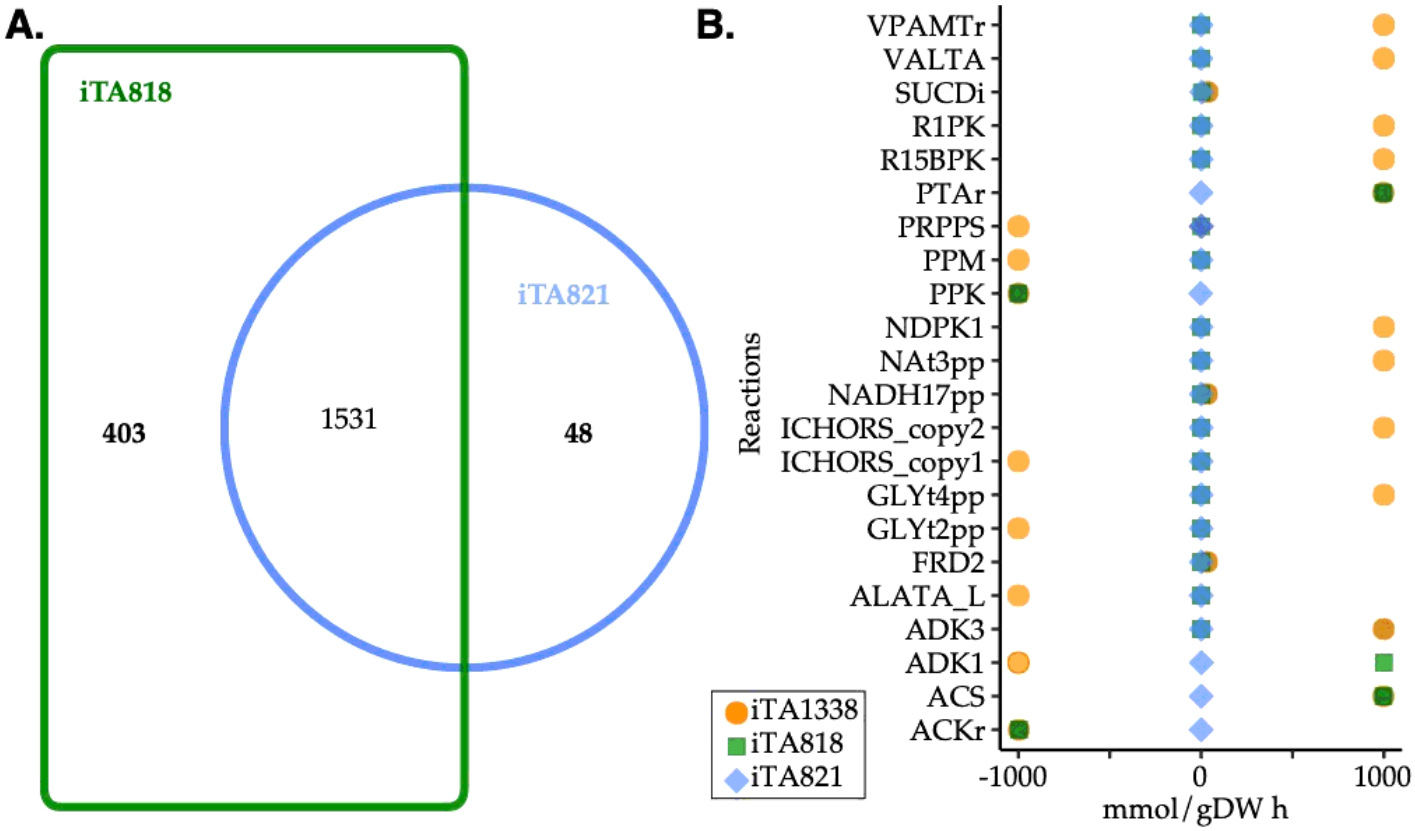

Figure 4

Flux balance analysis after transcriptomic integration using gene inactivity moderated by metabolism and expression (GIMME) under aerobic conditions.

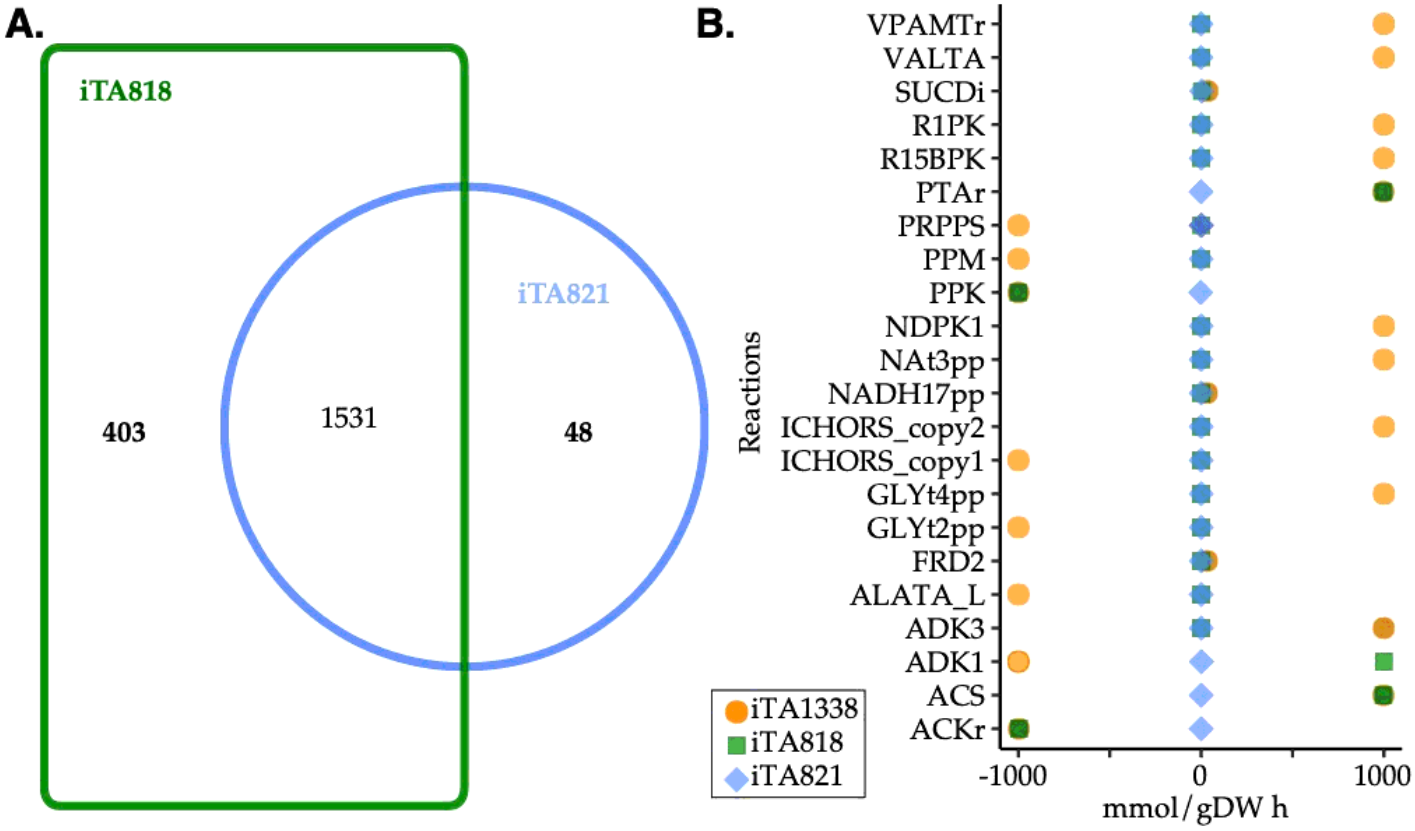

Figure 4

Flux balance analysis after transcriptomic integration using gene inactivity moderated by metabolism and expression (GIMME) under aerobic conditions.

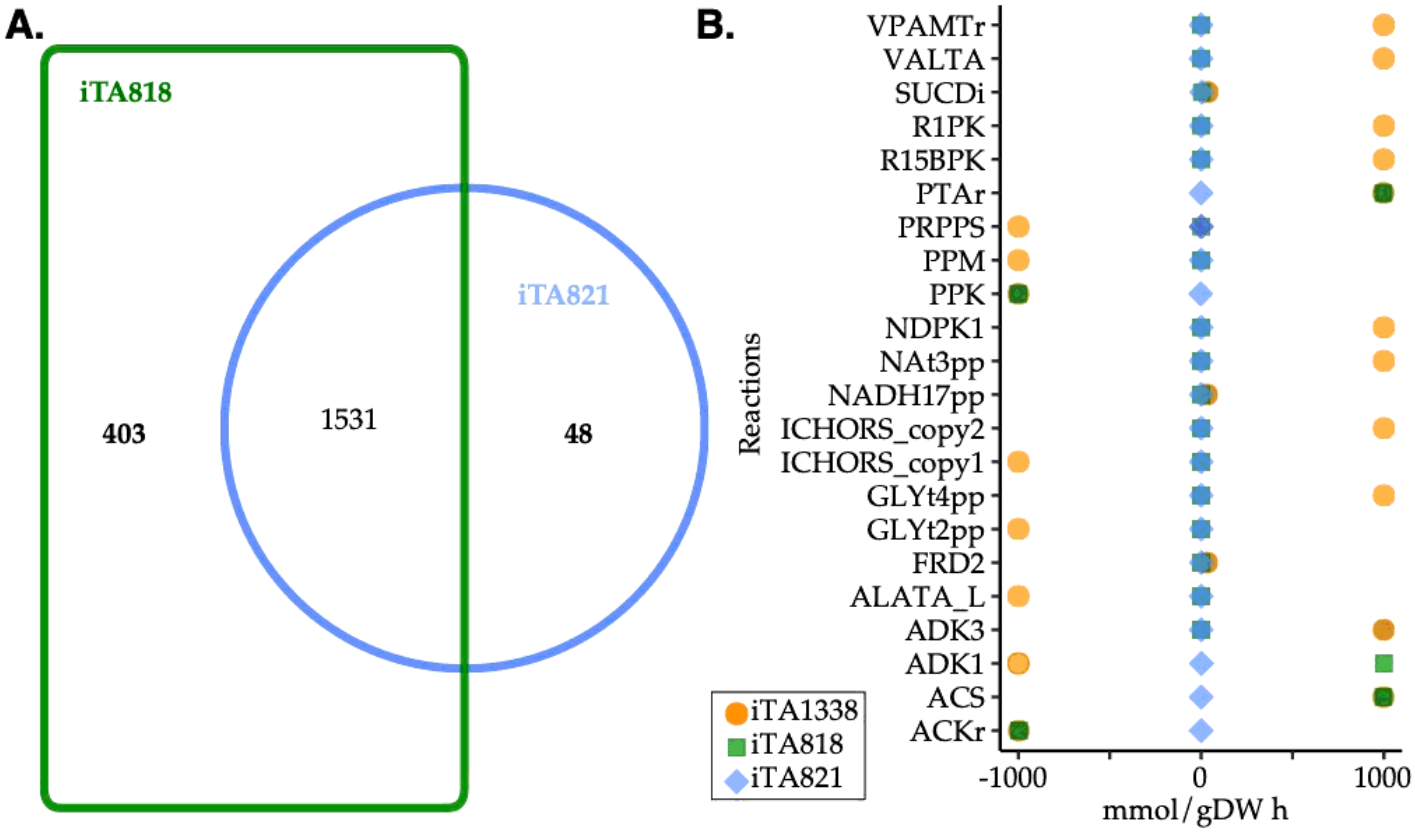

Figure 4

Flux balance analysis after transcriptomic integration using gene inactivity moderated by metabolism and expression (GIMME) under aerobic conditions.

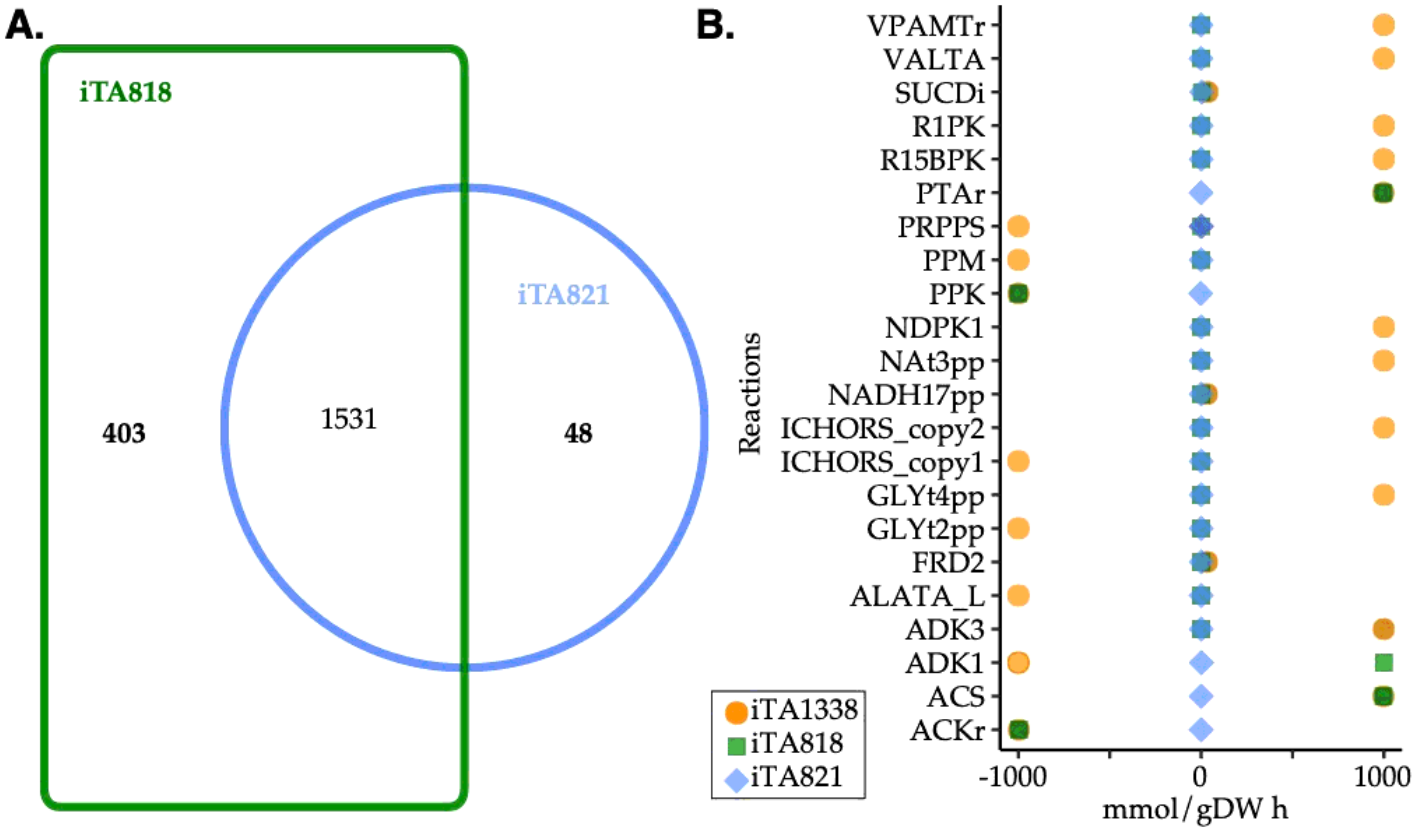

Figure 4

Flux balance analysis after transcriptomic integration using gene inactivity moderated by metabolism and expression (GIMME) under aerobic conditions.

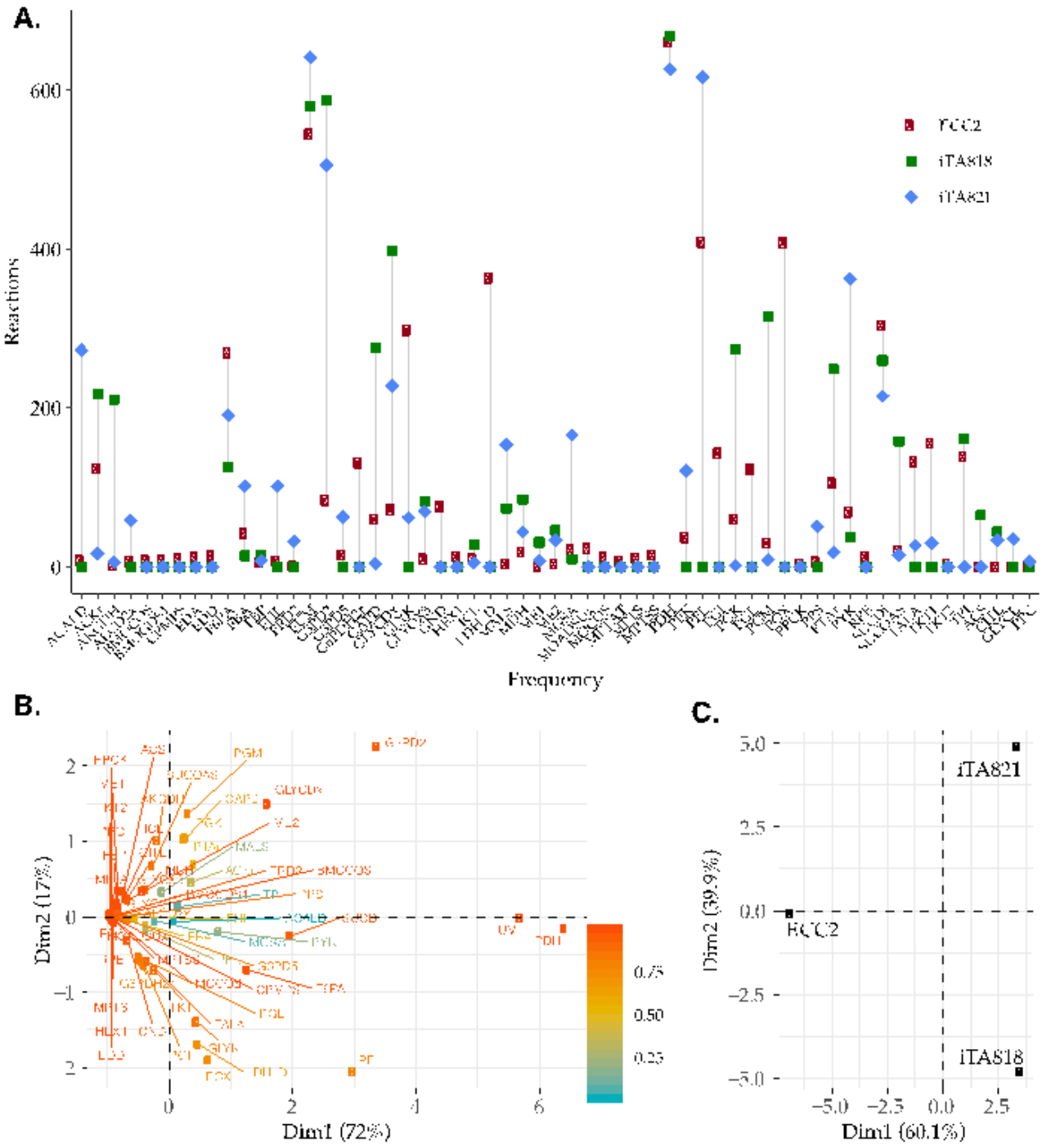

Figure 5

Frequency of reactions predicted by OptKnock, with combinations of knockouts from 1 to 10 reactions per mutant.

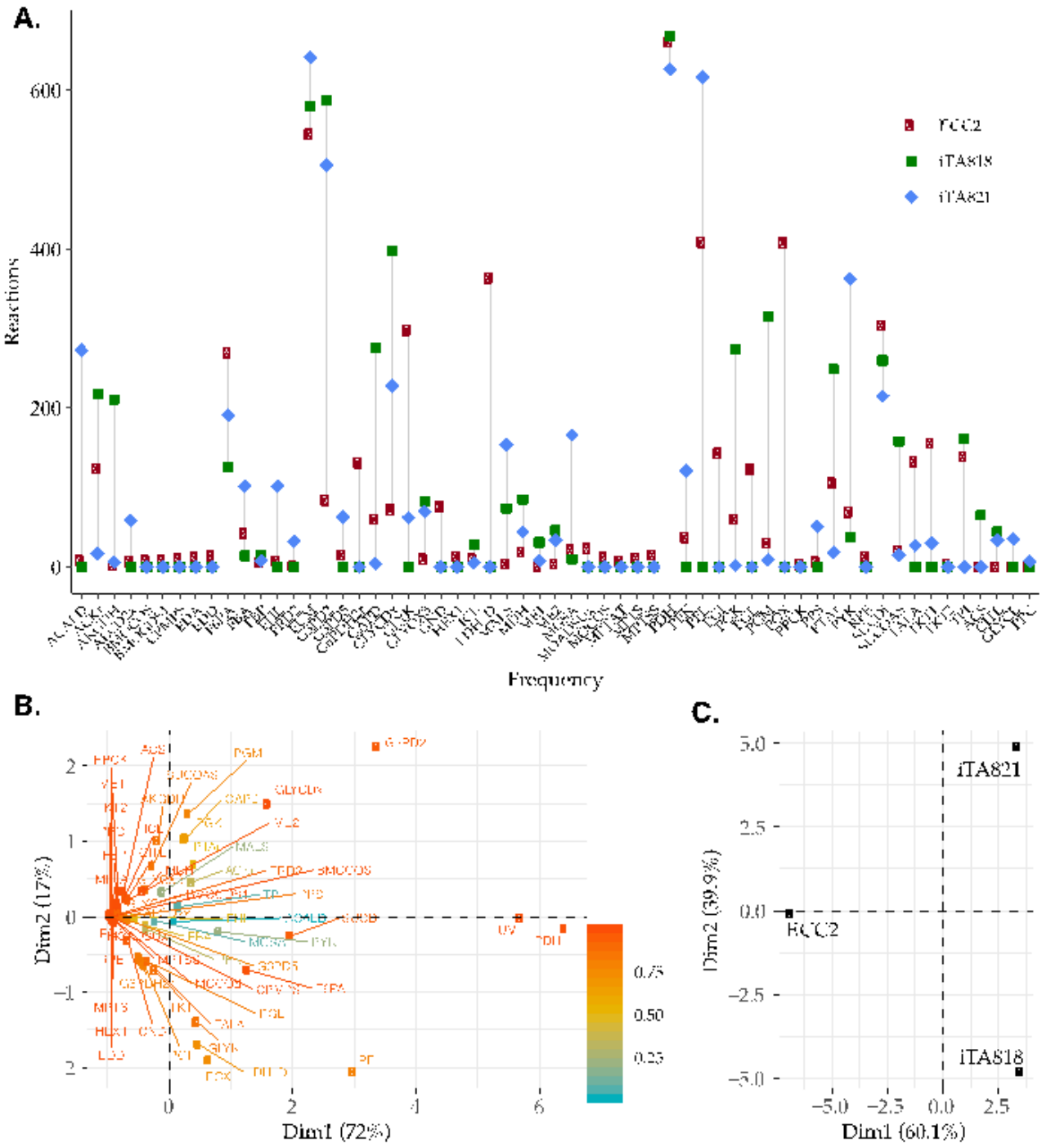

Figure 5

Frequency of reactions predicted by OptKnock, with combinations of knockouts from 1 to 10 reactions per mutant.

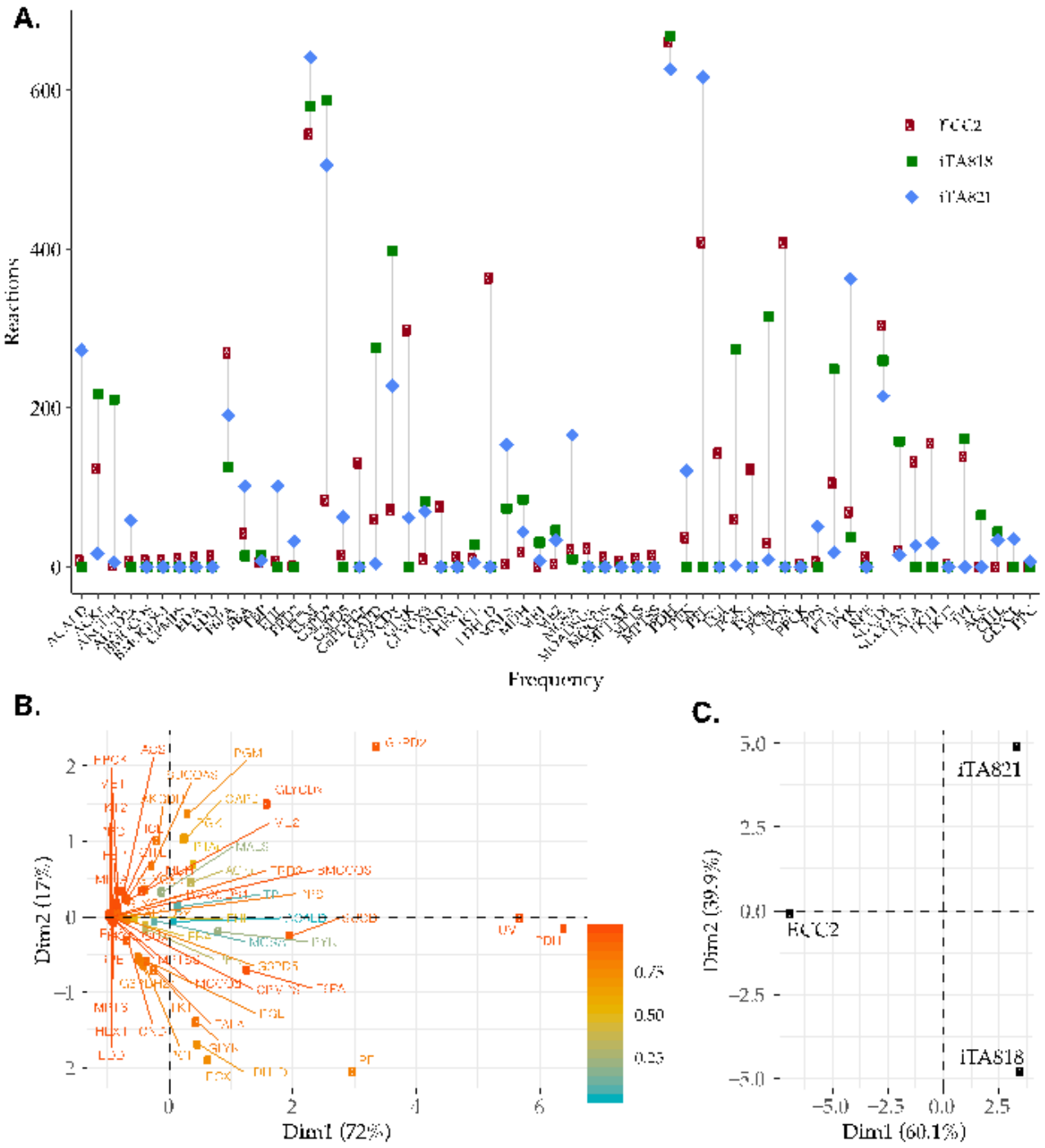

Figure 5

Frequency of reactions predicted by OptKnock, with combinations of knockouts from 1 to 10 reactions per mutant.

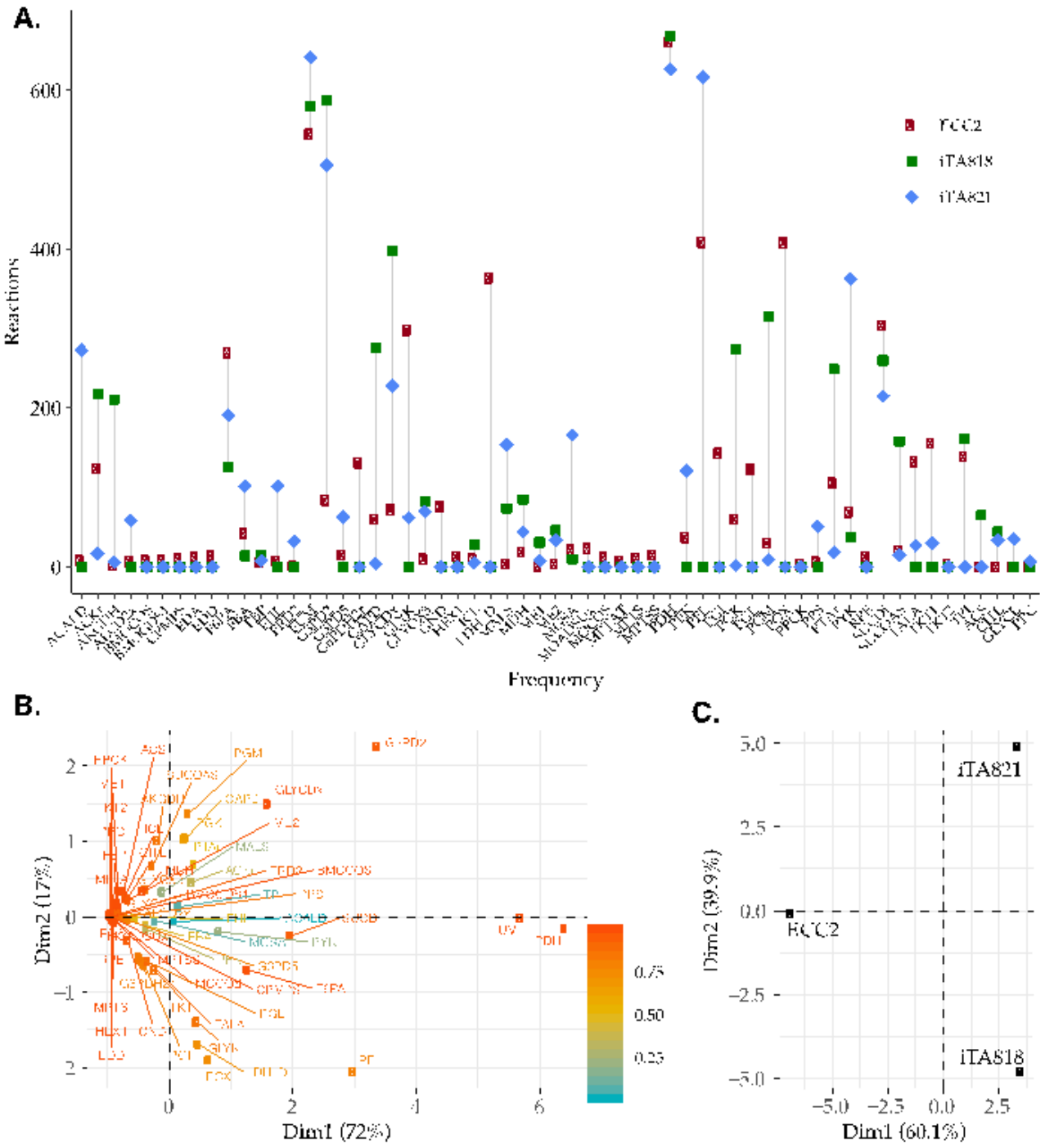

Figure 5

Frequency of reactions predicted by OptKnock, with combinations of knockouts from 1 to 10 reactions per mutant.

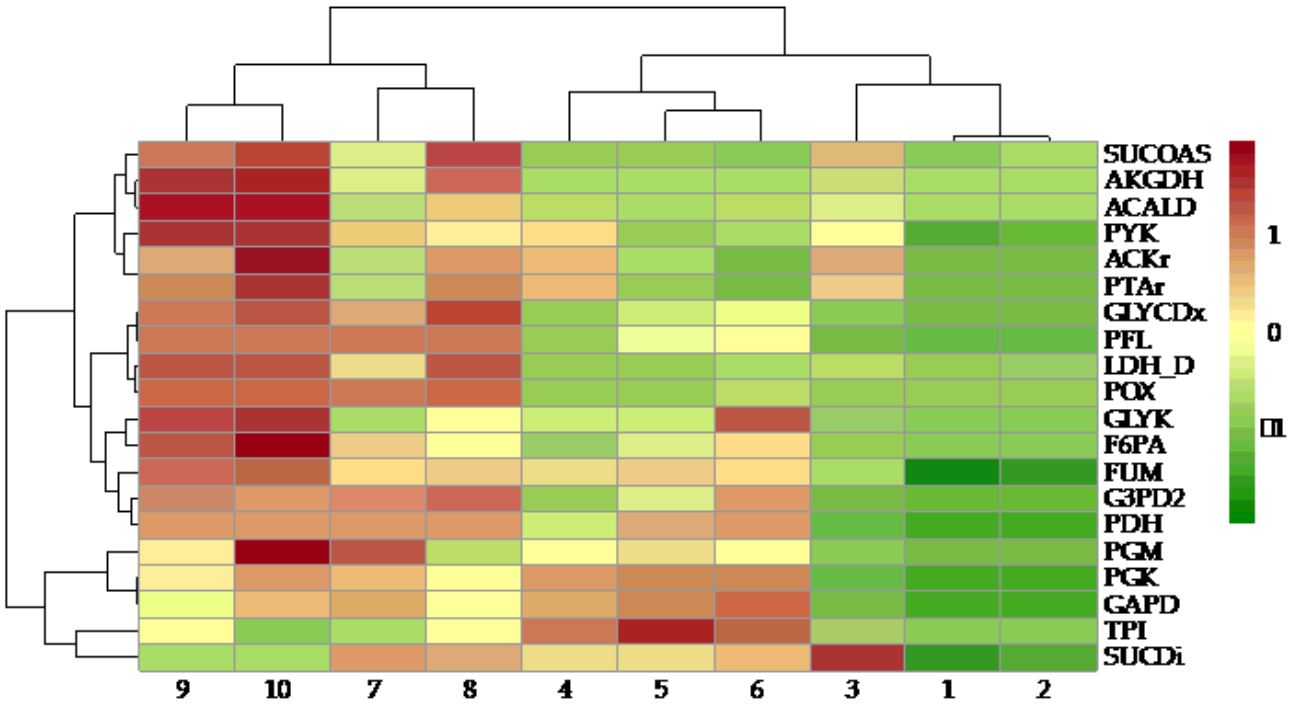

Figure 6

Pearson correlation of reaction frequency by knockouts numbers (k), predicted using OptKnock.

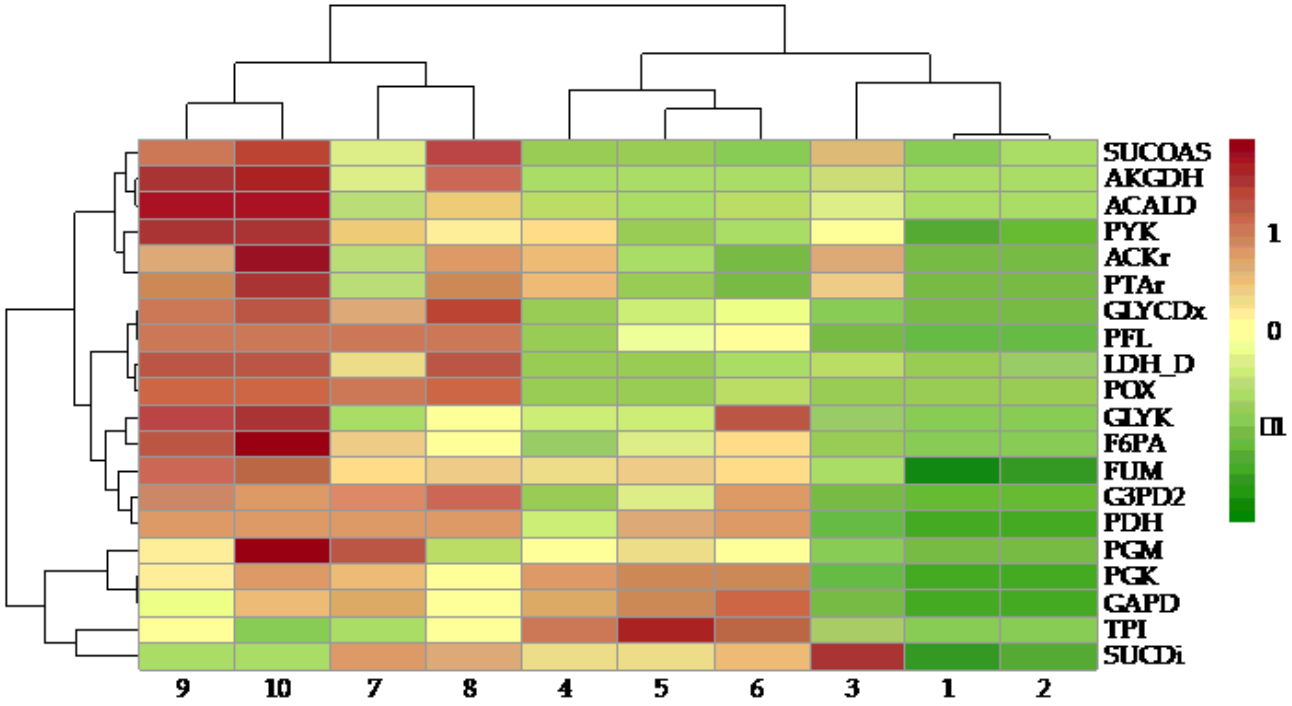

Figure 6

Pearson correlation of reaction frequency by knockouts numbers (k), predicted using OptKnock.

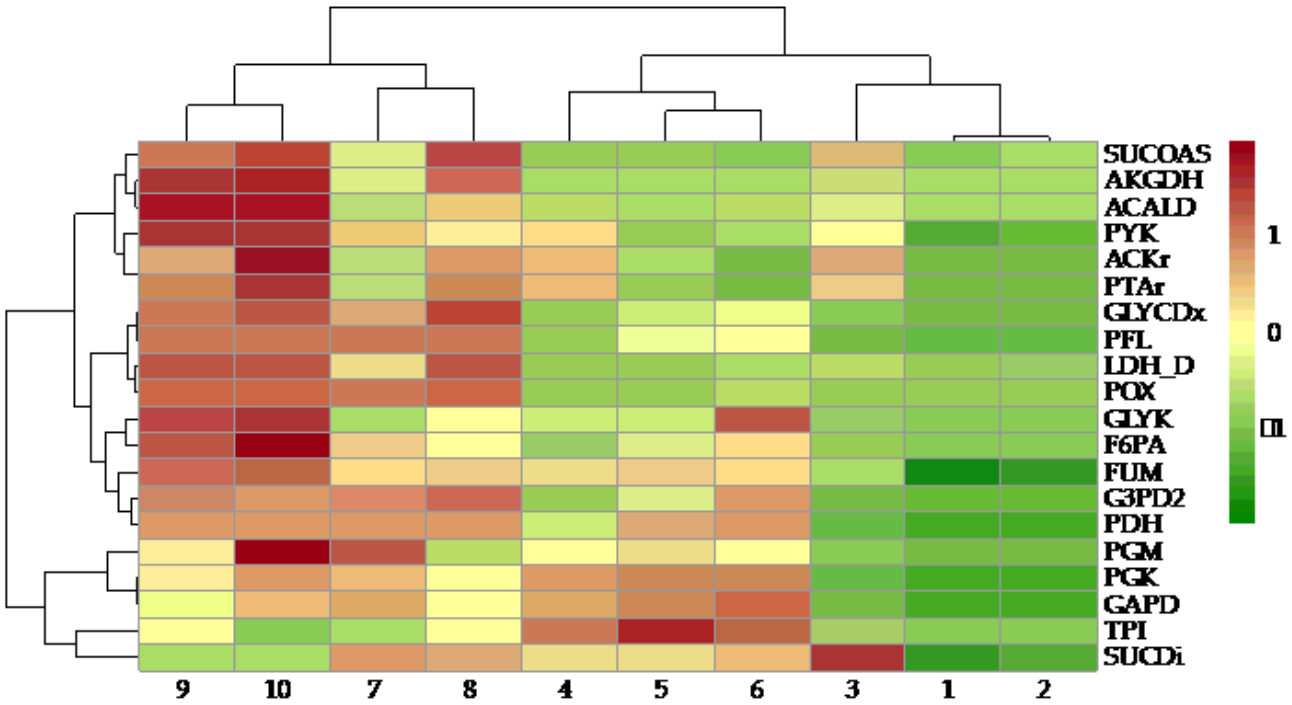

Figure 6

Pearson correlation of reaction frequency by knockouts numbers (k), predicted using OptKnock.

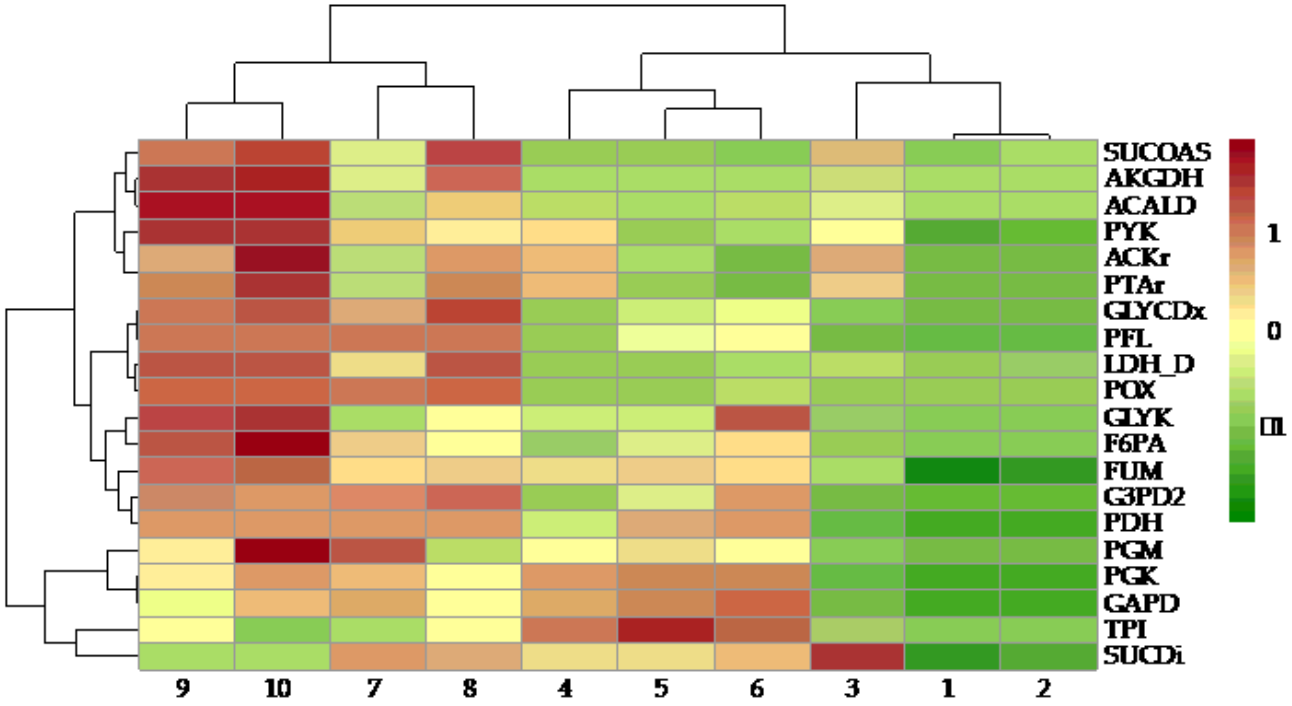

Figure 6

Pearson correlation of reaction frequency by knockouts numbers (k), predicted using OptKnock.

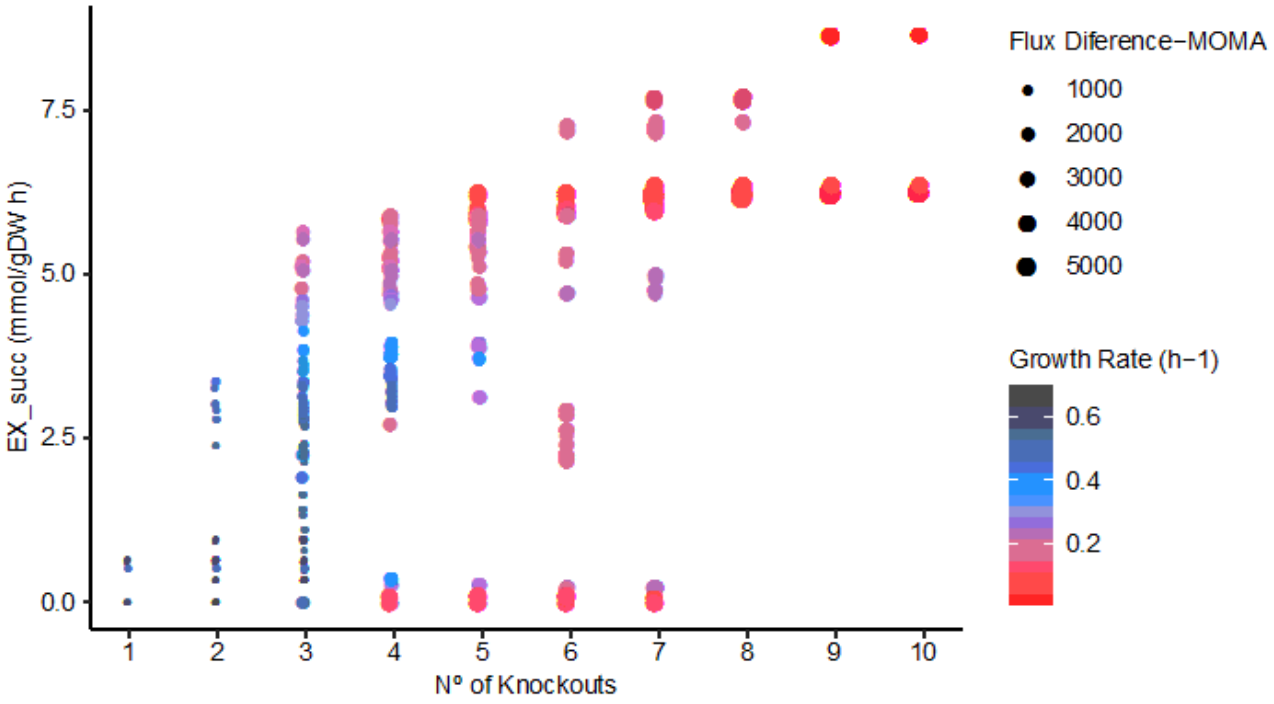

Figure 7

Relationship between the number of knockouts, succinic acid production, the growth rate, and the flux difference.

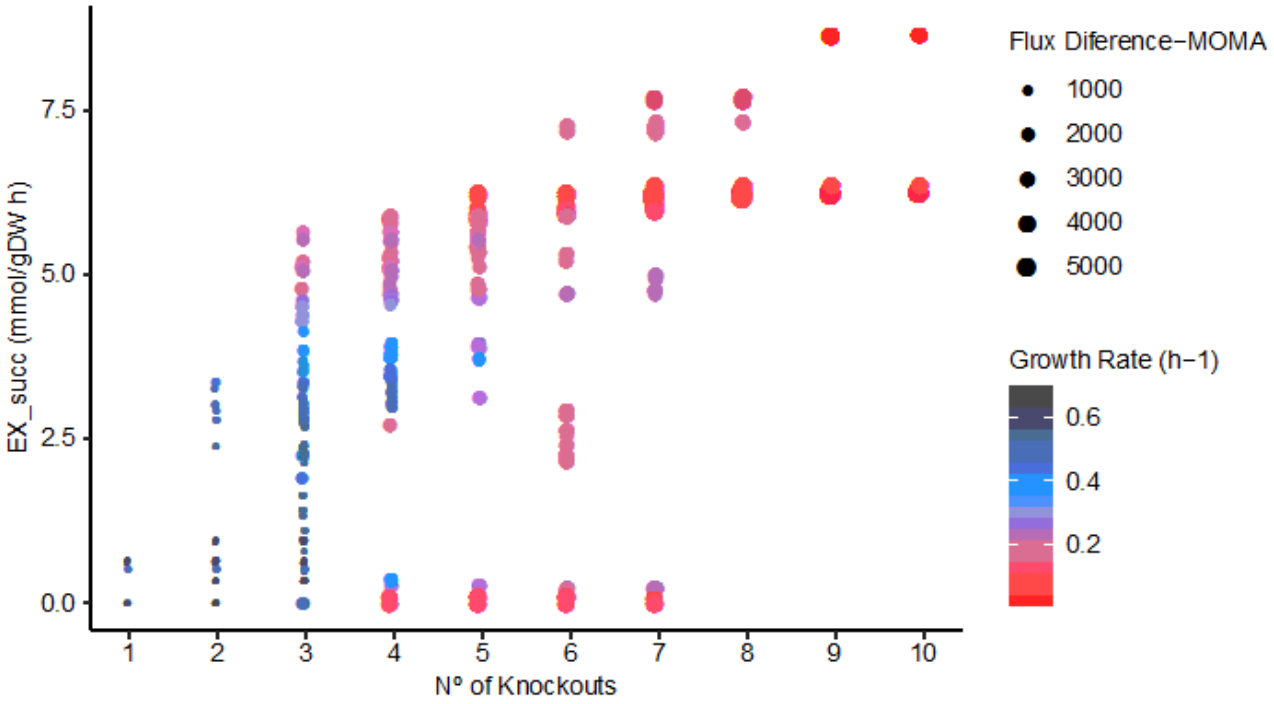

Figure 7

Relationship between the number of knockouts, succinic acid production, the growth rate, and the flux difference.

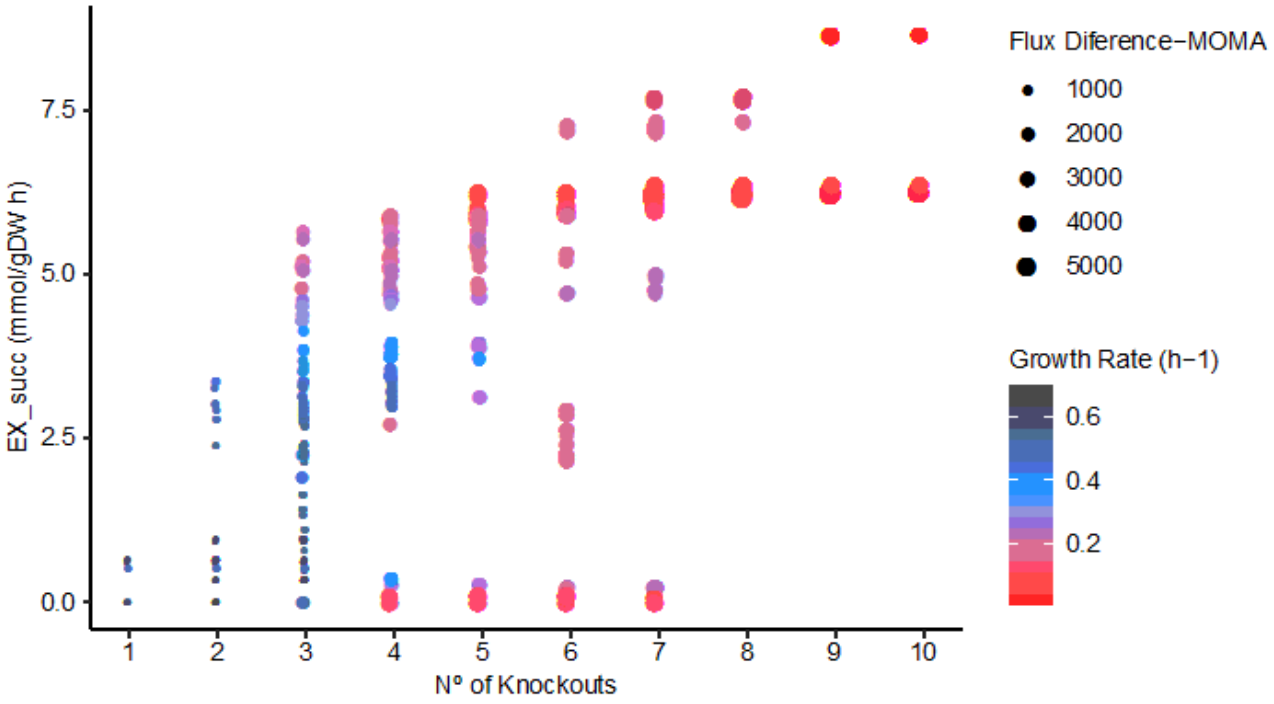

Figure 7

Relationship between the number of knockouts, succinic acid production, the growth rate, and the flux difference.

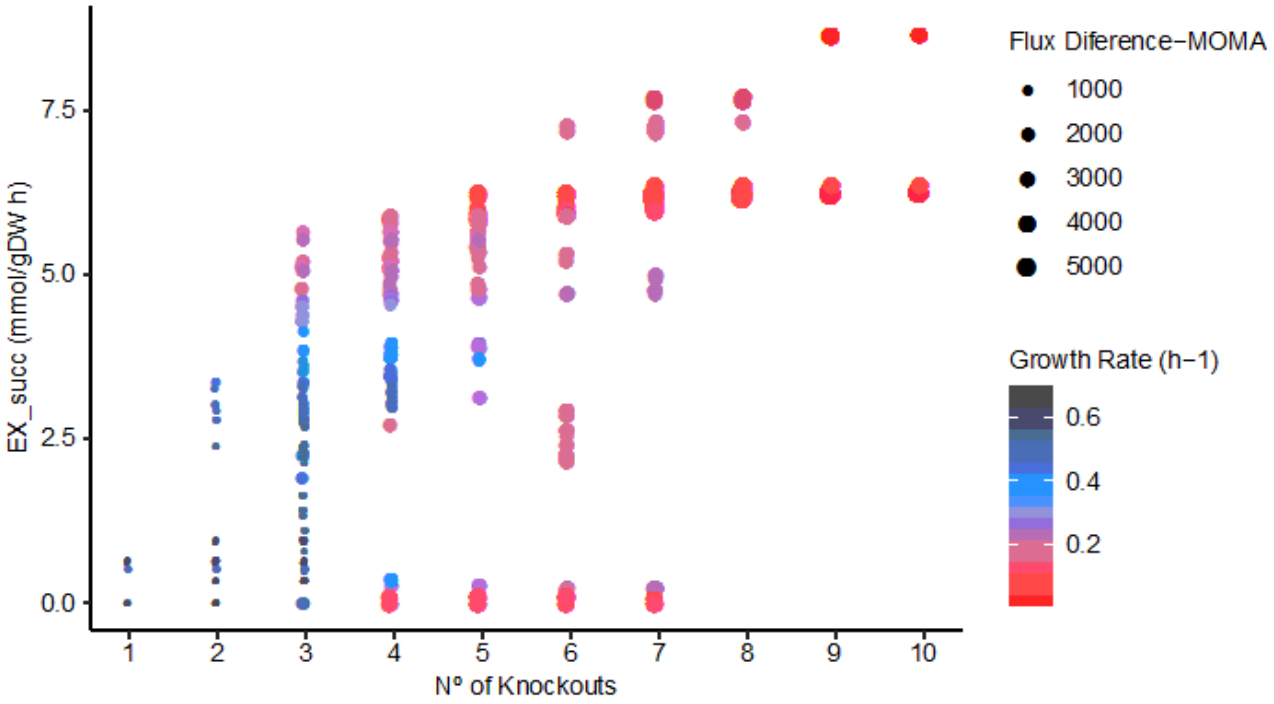

Figure 7

Relationship between the number of knockouts, succinic acid production, the growth rate, and the flux difference.

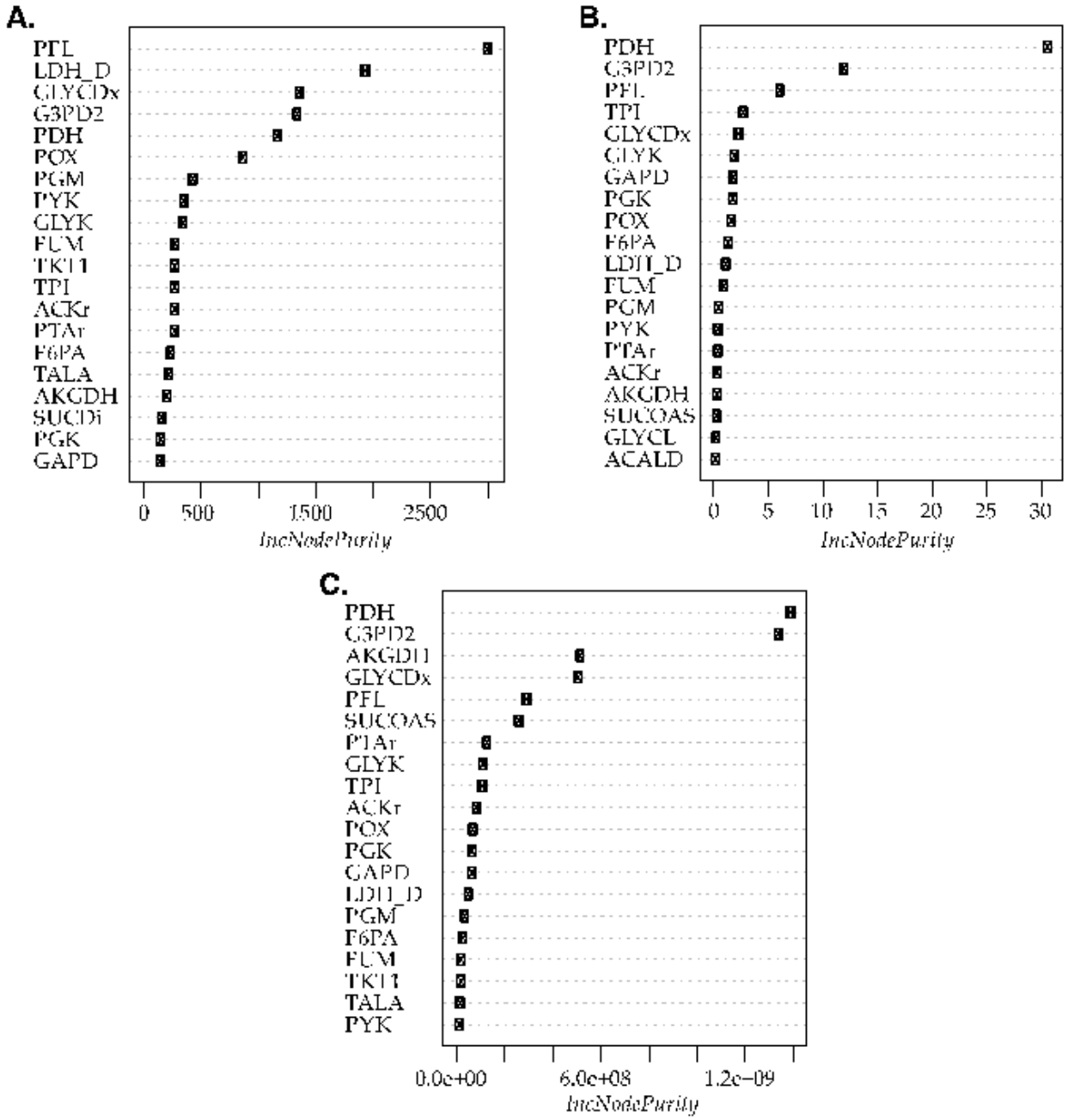

Figure 8

Top 20 node purities obtained using random forest for reactions predicted by OptKnock. A. Reaction knockout importance on succinic acid production. B. Reaction knockout importance in growth rate reduction. C. Reaction knockout importance in Euclidian distance.

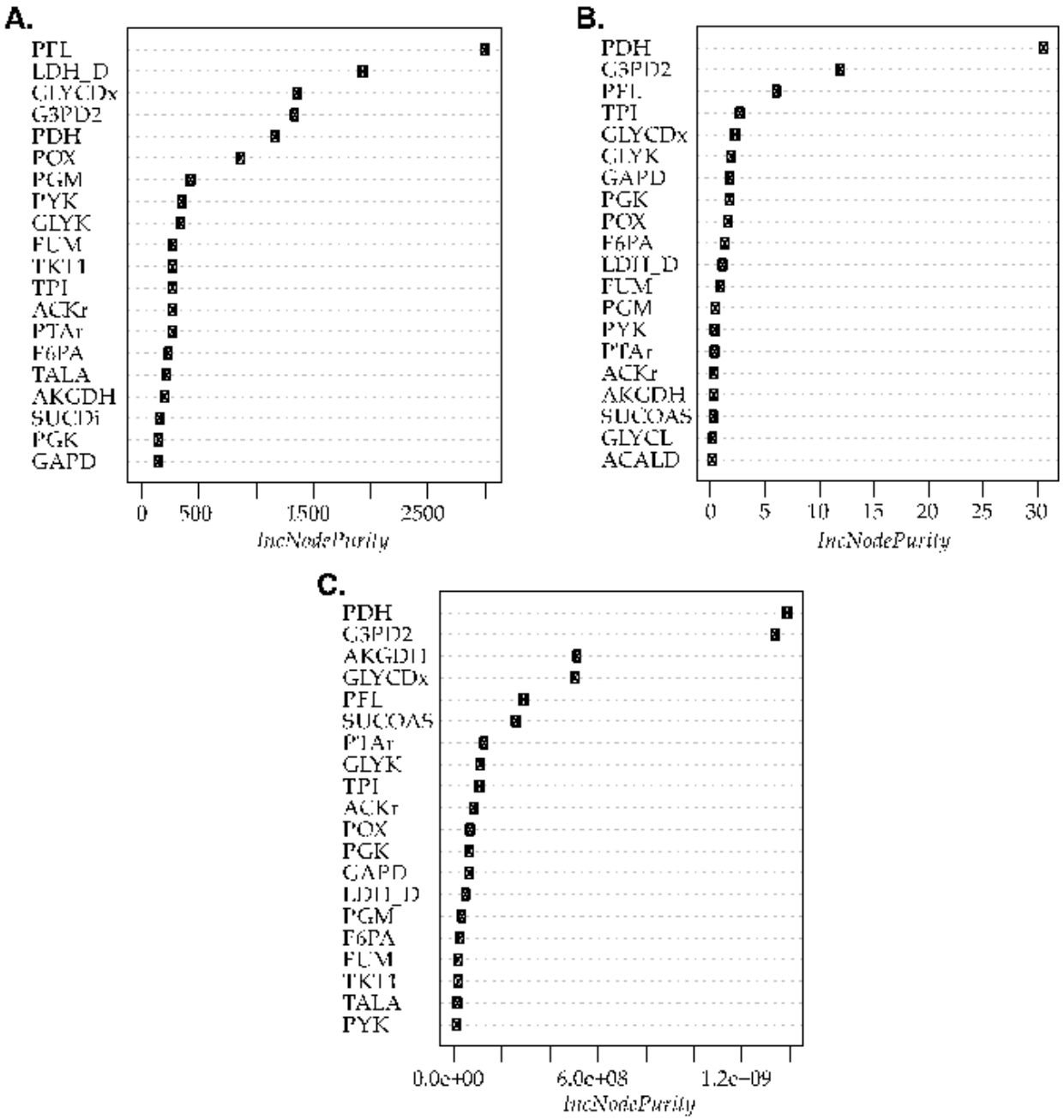

Figure 8

Top 20 node purities obtained using random forest for reactions predicted by OptKnock. A. Reaction knockout importance on succinic acid production. B. Reaction knockout importance in growth rate reduction. C. Reaction knockout importance in Euclidian distance.

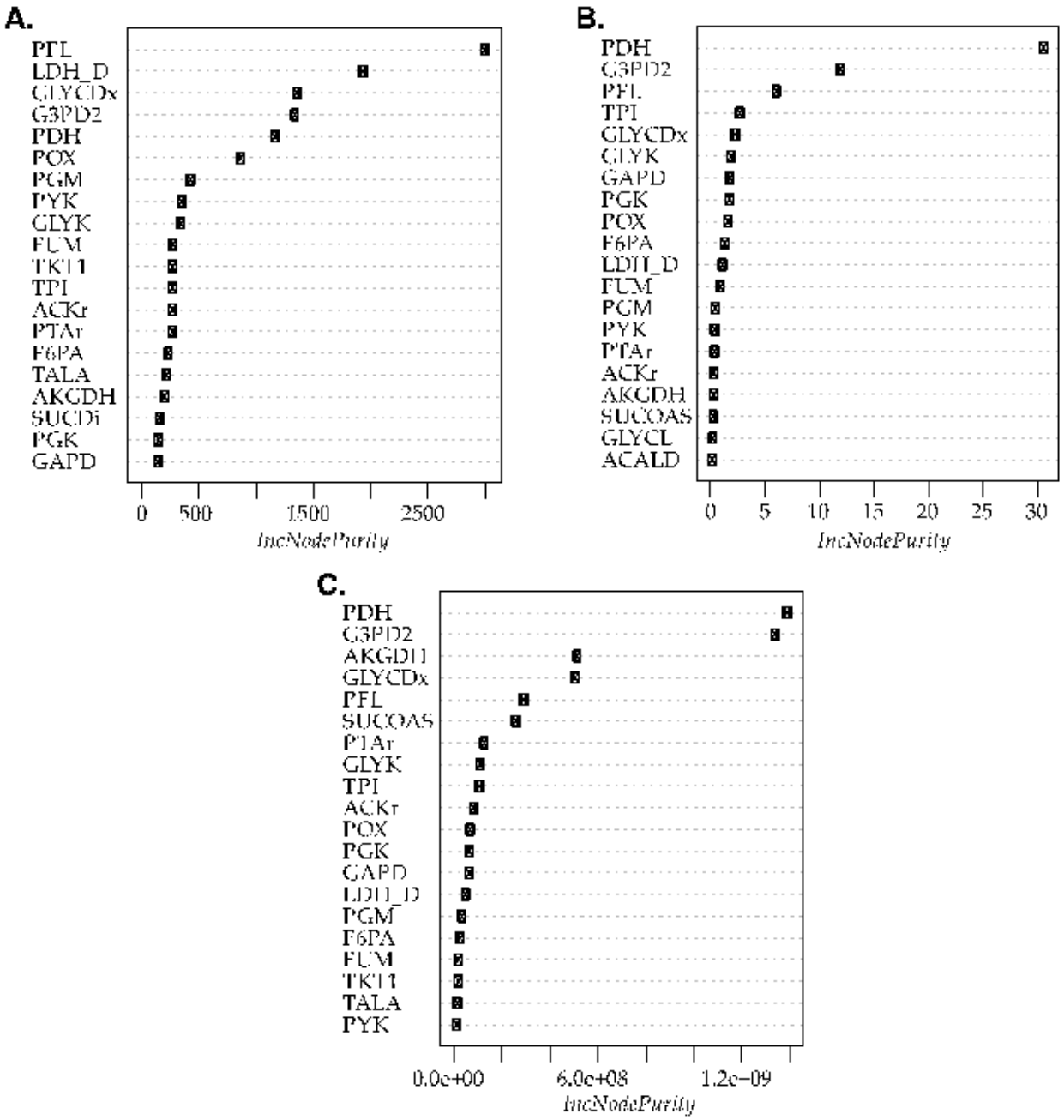

Figure 8

Top 20 node purities obtained using random forest for reactions predicted by OptKnock. A. Reaction knockout importance on succinic acid production. B. Reaction knockout importance in growth rate reduction. C. Reaction knockout importance in Euclidian distance.

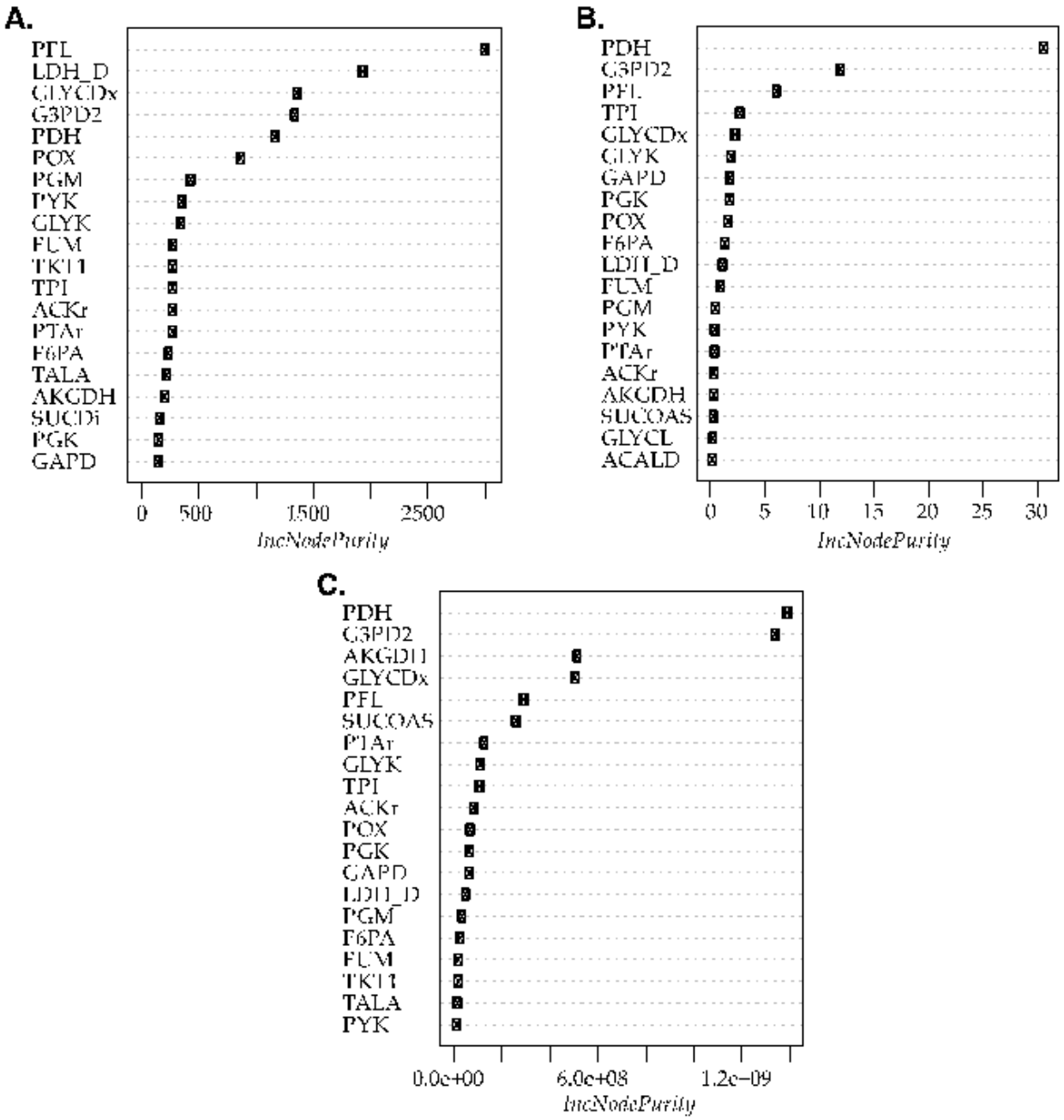

Figure 8

Top 20 node purities obtained using random forest for reactions predicted by OptKnock. A. Reaction knockout importance on succinic acid production. B. Reaction knockout importance in growth rate reduction. C. Reaction knockout importance in Euclidian distance.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryfileS1.xlsx
- SupplementaryfileS1.xlsx

- SupplementaryfileS1.xlsx
- SupplementaryfileS1.xlsx