In silico Design for Systems-Based Metabolic Engineering for the Bioconversion of Valuable Compounds From Industrial By-Products

Albert Enrique Tafur Rangel[1,2]*, Wendy Ríos[1], Daisy Mejía[1], Carmen Ojeda[1], Ross Carlson[3], Jorge Mario Gómez Ramírez[1] and Andrés Fernando González Barrios[1]*

[1] Grupo de Diseño de Productos y Procesos, Department of Chemical and Food Engineering, Universidad de los Andes, Bogotá, Colombia, [2] Grupo de Investigación CINBIOS, Department of Microbiology, Universidad Popular del Cesar, Valledupar, Colombia, [3] Center for Biofilm Engineering, Montana State University, Bozeman, MT, United States

Selecting appropriate metabolic engineering targets to build efficient cell factories maximizing the bioconversion of industrial by-products to valuable compounds taking into account time restrictions is a significant challenge in industrial biotechnology. Microbial metabolism engineering following a rational design has been widely studied. However, it is a cost-, time-, and laborious-intensive process because of the cell network complexity; thus, it is important to use tools that allow predicting gene deletions. An *in silico* experiment was performed to model and understand the metabolic engineering effects on the cell factory considering a second complexity level by transcriptomics data integration. In this study, a systems-based metabolic engineering target prediction was used to increase glycerol bioconversion to succinic acid based on *Escherichia coli*. Transcriptomics analysis suggests insights on how to increase cell glycerol utilization to further design efficient cell factories. Three *E. coli* models were used: a core model, a second model based on the integration of transcriptomics data obtained from growth in an optimized culture media, and a third one obtained after integration of transcriptomics data from adaptive laboratory evolution (ALE) experiments. A total of 2,402 strains were obtained with fumarase and pyruvate dehydrogenase being frequently predicted for all the models, suggesting these reactions as essential to increase succinic acid production. Finally, based on using flux balance analysis (FBA) 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 the importance of each knockout's (feature's) contribution. Glycerol has become an interesting carbon source for industrial processes due to biodiesel business growth since it has shown promising results in terms of biomass/substrate yields. The combination of transcriptome, systems metabolic modeling, and machine

learning analyses revealed the versatility of computational models to predict key metabolic engineering targets in a less cost-, time-, and laborious-intensive process. These data provide a platform to improve the prediction of metabolic engineering targets to design efficient cell factories. Our results may also work as a guide and platform for the selection/engineering of microorganisms for the production of interesting chemical compounds.

Keywords: systems metabolic engineering, transcriptomics, machine learning, adaptive laboratory evolution, metabolic modeling resources/frameworks

INTRODUCTION

Shifting from petrochemical sources to renewable, abundant, and inexpensive feedstocks to obtain valuable chemicals has become a promising goal for the chemical industry (Vlysidis et al., 2011). The biodiesel industry has increased in the last years by using renewable raw materials, but it generates large amounts of glycerol, which has become a burden. The bioconversion of glycerol is a potential route to increasing the use of bio-based succinic acid, a critical building block chemical with an attractive market. The availability of three pathways for succinic acid production (**Figure 1**; Chen et al., 2013a), 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 a 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, as compared with the petroleum-based approach.
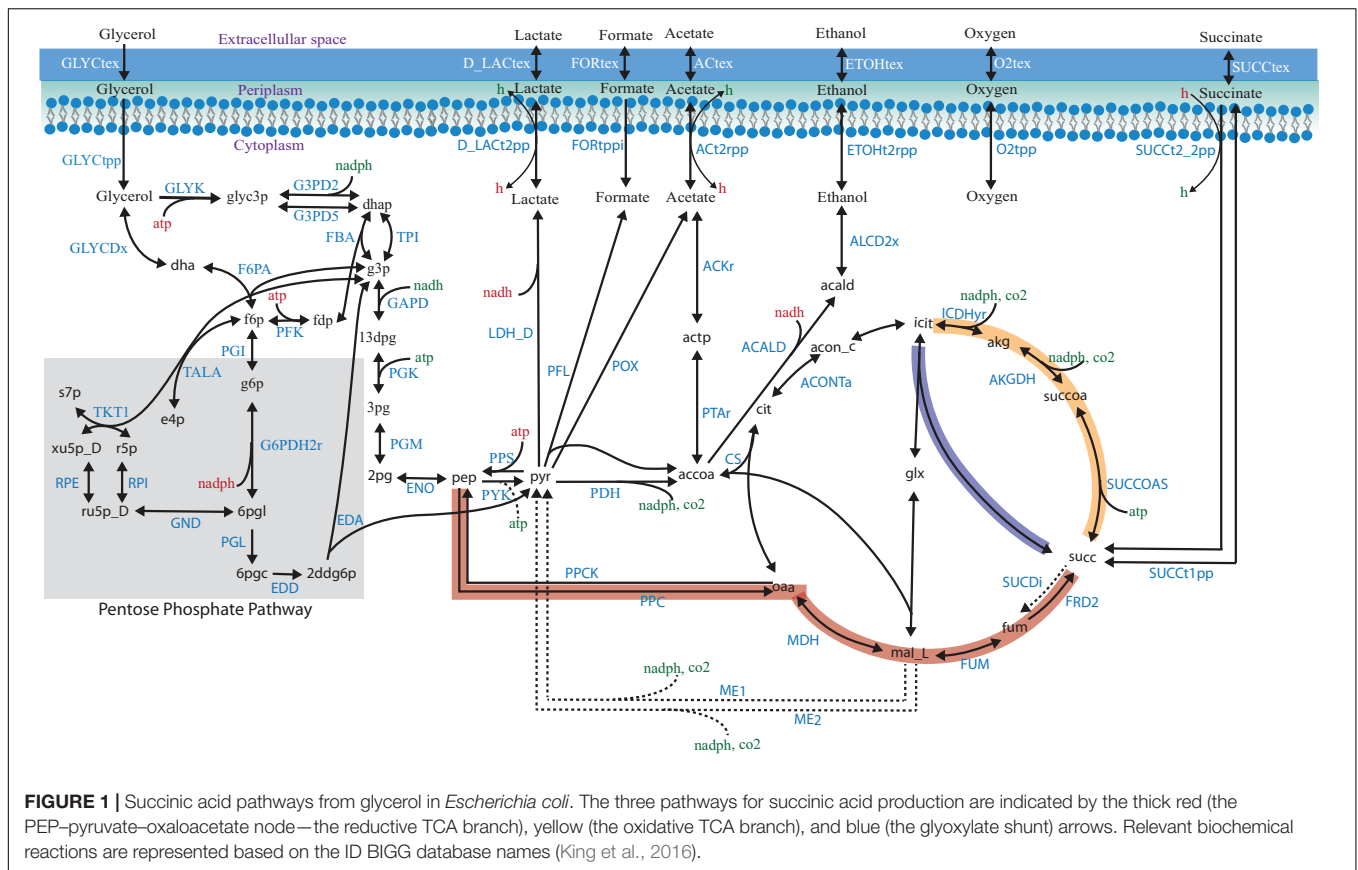
The main goal of using microbial cell factories is to design cheap and high-yield biotechnology-based conversion processes. A significant problem to be solved is how to enhance cell growth while using its capabilities to obtain a high-yield target chemical product. A classical approach for that is adaptive 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 selects targets that increase 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 (Kern A. et al., 2007; Chen et al., 2013a,b; Förster and Gescher, 2014; Woo and Park, 2014), including bio-based organic acids by substrate transport enhancement, gene overexpression, and deletion (Shams Yazdani and Gonzalez, 2008; Zhang B. et al., 2012; Buschke et al., 2013; Förster and Gescher, 2014; Yin et al., 2015; Zhu and Jackson, 2015). However, making the strain industrially competitive requires much time, effort, and high cost (Rangel et al., 2020).

When DNA was discovered in the last century, a new approach called metabolic network modeling for the study of cellular metabolism was developed (O'Brien et al., 2015). It allows to determine how several pathways in a cell can interact, as well as to elucidate basic microbial processes (Haggart et al., 2011). The first genome-scale metabolic network was described in 1999, and in 2002, the use of metabolic modeling to analyze recombinant pathways was reported (Carlson et al., 2002). Several models have been developed ever since with significant accuracy and useful predictions (Portela et al., 2013) that can be used to guide experimental studies (Pharkya and Maranas, 2006; O'Brien et al., 2015). COnstraints-Based Reconstruction and Analysis (COBRA) methods make it 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 (Kim, 2012; Ruckerbauer et al., 2015). 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 (Burgard et al., 2003; Pharkya et al., 2003; Pharkya et al., 2004; Pharkya and Maranas, 2006; Ranganathan et al., 2010).

OptKnock applies a flux balance analysis (FBA) approach for simulating genome-scale metabolic models (GEMs). 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) were developed (Segre et al., 2002). 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 a more accurate prediction of the immediate metabolic response to KO than FBA does on its own. However, a large list of knockout combinations could be obtained when computational tools are used, and select which test in a lab can be laborious.

Several approaches to optimize cell factories have been developed, but conventional and computational approaches

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 (King et al., 2016).

have not always been successful due to unexpected changes in the cell where an intracellular complex interconnected network of genes, proteins, and reactions exists. Systems metabolic engineering has emerged as an approach that integrates metabolic engineering and combined metabolic and "omics" network models. This approach could be beneficial for genome-scale modeling because it reduces the solution space and generates accurate predictions (Nordlander et al., 2008; Feist et al., 2010; Blazier and Papin, 2012; Machado and Herrgård, 2014; Rangel et al., 2020). Mainly considering that under certain environmental conditions, there are a limited number of reactions that are active according to transcriptional responses and other regulation phenomena to provide beneficial improvements for the cell bioconversion process (Fong and Marciniak, 2003; Fong et al., 2005; Lee and Palsson, 2010; Conrad et al., 2011; Wang et al., 2011; Zhang J. et al., 2012; Bao et al., 2014).

In this study, systems metabolic engineering for overproduction of succinic acid from glycerol in *E. coli* ATCC 8739 was used through integration of transcriptomics data to metabolic models and classification tree analysis using the 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 model 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 succinic acid production, growth rate, and Euclidian distance between the WT strain and each mutant as response variables.

MATERIALS AND METHODS

Strains and Culture Conditions

E. coli ATCC 8739 was used in this study. It was obtained commercially 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$_2$O, 0.17 g/L ZnCl$_2$, 0.043 g/L CuCl$_2$ 2H$_2$O, 0.06 g/L CoCl$_2$ 6H$_2$O, 0.06 g/L Na$_2$MoO$_4$ 2H$_2$O, 0.06 g/L Fe(NH$_4$)$_2$(SO4)$_2$ 6H$_2$O, 0.20 g/L FeCl$_3$ 6H$_2$O], 5 ml of MgSO$_4$ (1 M), and 30 g of glycerol. A 50 ml culture was carried out in a 250 ml baffled-conical Erlenmeyer flask and cultivated aerobically at 37°C and 200 rpm.

Two conditions were evaluated in this study: an adapted *E. coli* on high glycerol concentrations (30, 40, 50, 60 g/L) and an optimized culture condition. For the first condition, four *E. coli* cultures were continuously subcultured each for 72 h in Luria–Bertani (LB) medium (5 g/L of NaCl, 10 g/L of tryptone, and

5 g/L of yeast extract, supplemented with 30, 40, 50, or 60 g/L). After every three subcultured rounds (216 h), the concentration of tryptone was decreased from 1 until reaching 0 g/L. Then, 10 subcultured rounds each for 72 h were carried out. During the complete experiment, 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. For each subcultured round, an OD ~0.33 600 nm was considered as inoculum starting point. At the end of each tryptone decreasing, 1 ml of culture was kept at −80°C and used for further evaluation of growth and glycerol uptake. 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 at the same conditions as the reference culture.

Differential Expression Analysis

RNA-Seq was carried out in triplicate for all conditions. For the adapted strain, the culture conditions for RNA-Seq were the same as those for the optimized culture medium condition. To harvest cells for total RNA purification, the 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 to read counts. SARTools (Statistical Analysis of RNA-Seq data Tools) (Varet et al., 2016) 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 Gene Ontology (GO) analysis by Blast2GO (Götz et al., 2008). The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar, 2002) 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, two *E. coli* models were used: *EColiCore2* (ECC2) (data under peer review) and iTA1338 for *E. coli* ATCC 8739 (**Supplementary File 1**). 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. GapFind and GapFill, 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 1**). All optimization problems were solved using the COBRA Toolbox v.3.0 (Heirendt et al., 2019).

Transcriptomics Integration and Metabolic Engineering Target Prediction

The gene inactivity moderated by metabolism and expression (GIMME) (Becker and Palsson, 2008) 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 log2 fold change (FC) $\geq |1|$. Next, according to the GPR 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. However, MOMA was used to understand the probability of those mutants predicted to be adapted and to reach the optimal state (predicted succinic and growth flux) considering the Euclidean distance. It because OptKnock predicts an optimal state, but after genetic manipulation cell are not in this state. The maximum uptake rate of glycerol was set to 13.3 mmol/g DW h^{-1}. The OptKnock, GIMME, and MOMA methods were conducted using COBRA Toolbox v.3.0 (Heirendt et al., 2019) 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 (Li et al., 2015). For interpretation purposes, this measure can be used to rank variables by the strength of their relation to the response variable (Li et al., 2015). 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 the 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.

RESULTS

Glycerol Consumption of *E. coli* After Adaptive Laboratory Evolution

Luria–Bertani is one of the most common cultures used industrially for the growth of *E. coli*. In order to increase glycerol consumption by *E. coli* on LB media, an ALE experiment was carried out. Results obtained in this study, before the ALE experiments, suggest that even when high cell density cultures are reached, a low consumption of glycerol is observed. For all the four conditions (supplementation of 30, 40, 50, or 60 g/L of glycerol), a growth curve was carried out, showing that a maximum of 7 g/L consumption of glycerol could be achieved naturally by *E. coli*. Nevertheless, after the ALE experiments, an increase of 3 g/L in the glycerol consumption was observed for the strain growing in a supplementation of 30 g/L of glycerol. Despite this data showing an increase of around 30% in glycerol consumption, it is far below that obtained in the optimized culture, which reaches a consumption of 30 g/L of glycerol (data under peer review).

Transcriptional Response of *E. coli* for Aerobic Glycerol Consumption

A cell is considered a complex system where a large number of processes are carried out. These processes then involve an interaction between genes, transcripts, proteins, metabolites, and reactions, among others (Lee et al., 2012; Furusawa et al., 2013; Rangel et al., 2020). Metabolic models are reconstructed by using genome information; however, it is well known that metabolism is given by environmental conditions by passing through a cell regulation process. This causes some genes to be turned on and off under certain conditions. To determine which reactions are active to obtain high accurate models, two transcriptomic profiles were obtained from an ALE experiment and an optimized culture medium.

DEGs were determined using the DESeq2 statistical package after filtering out low count reads with an average value of <100. Significant 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 $P < 0.05$. 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 $\geq$ |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 downregulated and 256 upregulated for the optimized medium, and 431 DEGs for the evolved strain, of which 223 genes were downregulated and 208 genes were upregulated. When comparing DEGs in the optimized medium and those in the evolved strain, 59 downregulated genes were found to be unique in the evolved strain and 58 unique genes for the optimized medium. In this context, 47 and 95 upregulated genes were found to be unique in the evolved strain and the optimized medium, respectively (**Figure 2**).

DEGs were classified into the following three groups using GO analysis: biological processes, molecular functions, and cellular components. The shared downregulated genes predominantly included those involved in the metabolic process (cellular, organic substances, nitrogen compounds, and primary metabolic processes), chemicals, stress and stimulus responses, and heterocyclic compound systems. Between downregulated genes, we found *phoB* and *phoR*, which are involved in phosphorous uptake and metabolism since, under excess phosphorous, PhoR inactivates *phoB* (Makino et al., 1989). **Figure 3** shows the level 2 GO terms for unique down- and upregulated genes in both conditions using Blast2GO (Götz et al., 2008). The 117 unique downregulated genes at log2 FC $\geq$ |2| and an adjusted $P \leq 0.05$ were classified into 15 functional groups. Two GO terms, signaling and locomotion, were only present for the evolved strain, and one GO term, multiorganism processes, was only present for the optimized culture condition in downregulated genes (**Figure 3A**).

GO analysis revealed that shared upregulated DEGs (**Figure 2**) are involved mostly in the metabolic process (51%), including GO terms such as cellular, organic substances, primary, and nitrogen compound processes; 11% of the upregulated genes were associated with biosynthetic processes and the establishment of localization. The main GO terms for molecular functions were those involved in a 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 upregulation 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 (Booth, 2014). Overexpression of *glpX* has been shown to increase hydrogen production (Kim et al., 2011). Additionally, transcriptomic analysis showed upregulation of both flavin oxidases *glpD* and *glpABC*.

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 also employ fumarate under anaerobic conditions by encoding a fumarate reductase complex under anaerobic
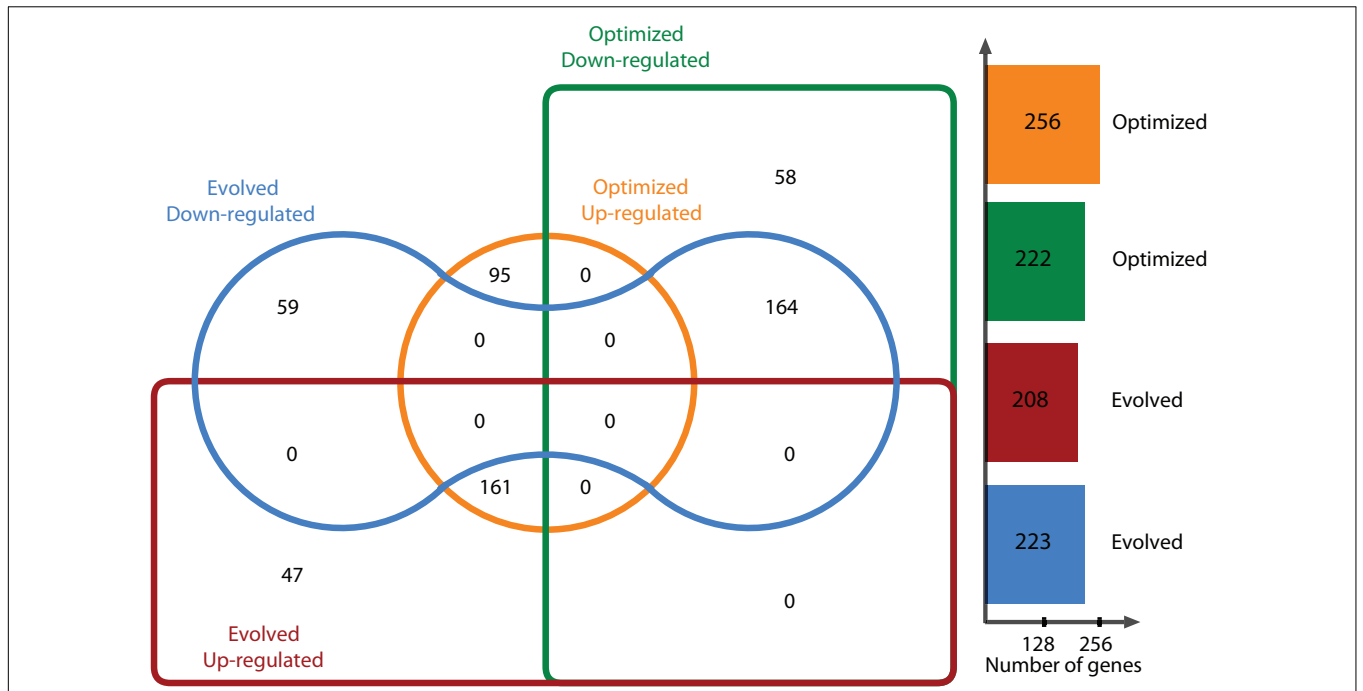
FIGURE 2 | *E. coli* ATCC 8739 differentially expressed genes in the optimized medium vs. glycerol-based medium and the evolved strain vs. glycerol-based medium.
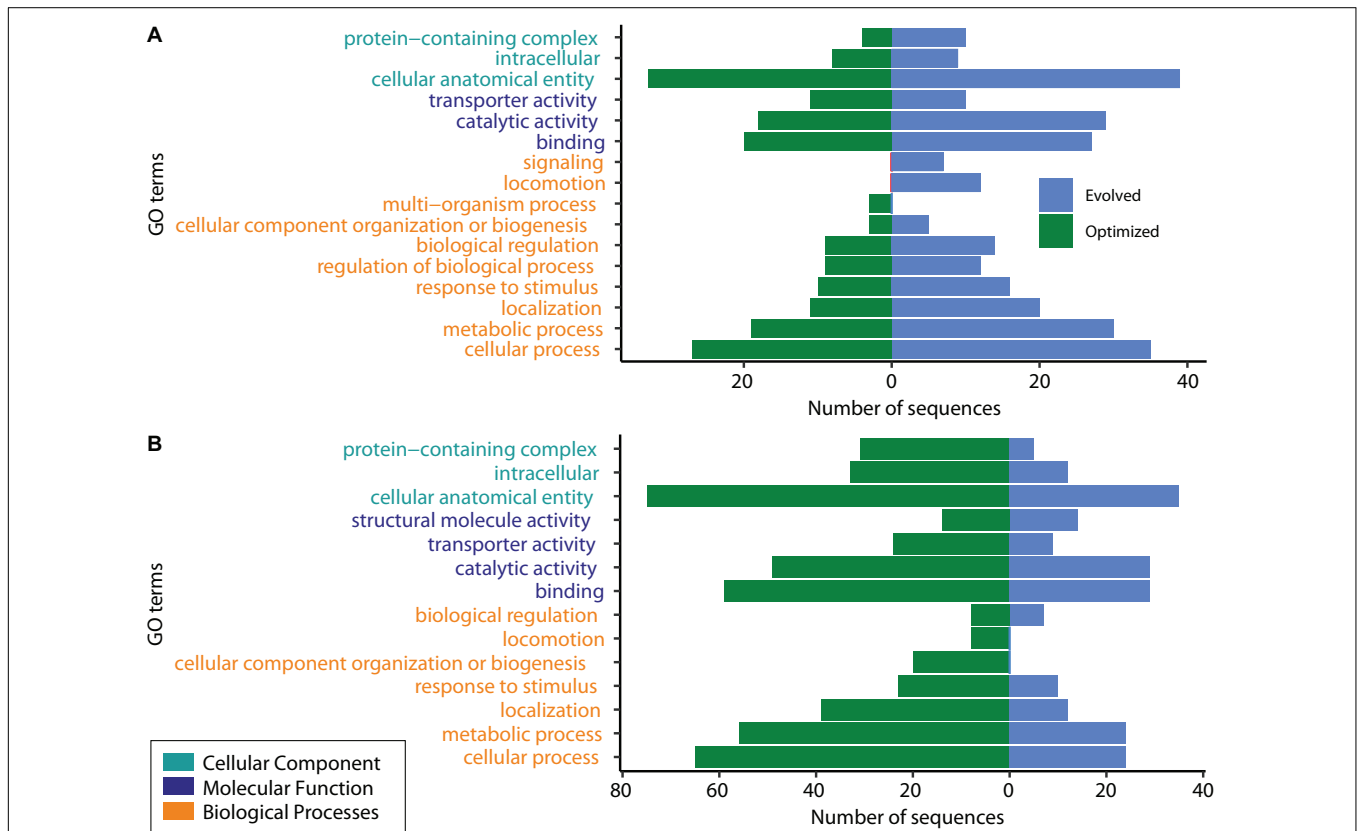
FIGURE 3 | Gene Ontology (GO) analysis of unique differentially expressed genes. **(A)** Downregulated genes. **(B)** Upregulated genes.

TABLE 1 | Differential expression of genes involved in glycerol metabolism.

RefSeq tag (ECOLC_RS)	Gene name	Old locus tag	Product	Log2 FC Exp 3	Log2 FC 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

conditions (Jones and Gunsalus, 1987; Cecchini et al., 2002). **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 has been reported to be predominantly expressed under aerobic conditions (Chen et al., 2012). Under aerobic conditions, the catalysis of succinate to fumarate interconversion is mediated by the succinate dehydrogenase complex encoded by *sdhABCD* (Cecchini et al., 2002). However, in this study, *sdhABCD* genes were not found to be differentially expressed in any of the culture conditions. Interestingly, among the 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 upregulation under glycerol utilization as a carbon source, and changes in the level of *glpK* transcription had a significant effect on *malT* transcription (Chagneau et al., 2001). In this study, *malEFKMTPQ* genes were shown to be upregulated 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, a 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 (Lin et al., 2005; Blankschien et al., 2010), was found to be upregulated (∼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 upregulated, 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 models (GEMs) 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 unexploited mainly (Ataman et al., 2017). To avoid this situation and to evaluate the effects of using 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 (Monk et al., 2013), iEC1349_Crooks (Monk et al., 2016), and iML1515 models (Monk et al., 2017). 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 2,032 metabolites, 2,804 reactions, and 1,338 genes (**Supplementary File 1**). After that, using GIMME, context-specific metabolic networks were constructed 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 1**) and (2) *E. coli*

ATCC 8739 strains evolved to grow on glycerol (iTA821) (**Supplementary File 1**). Manual curation was carried for the 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 (**Figure 4B**). The reactions only present in iTA821 are mainly associated with the inner membrane transport (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. Unusual 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.
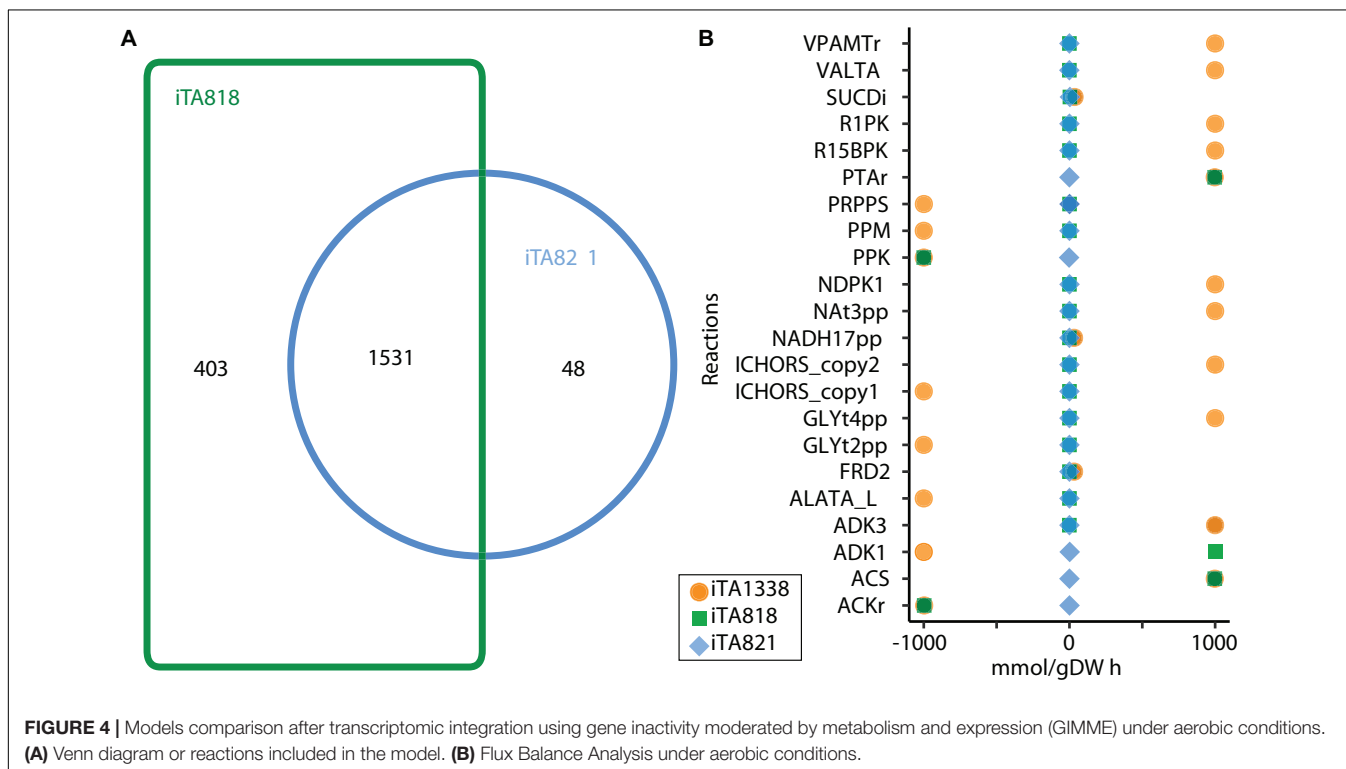
In silico Systems Metabolic Engineering Targets Prediction

To predict *E. coli* strains that overproduce succinic acid from glycerol, OptKnock was used (Burgard et al., 2003). Before predicting the reaction target to overproduce succinic acid, both metabolic networks were preprocessed. The goal of preprocessing 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
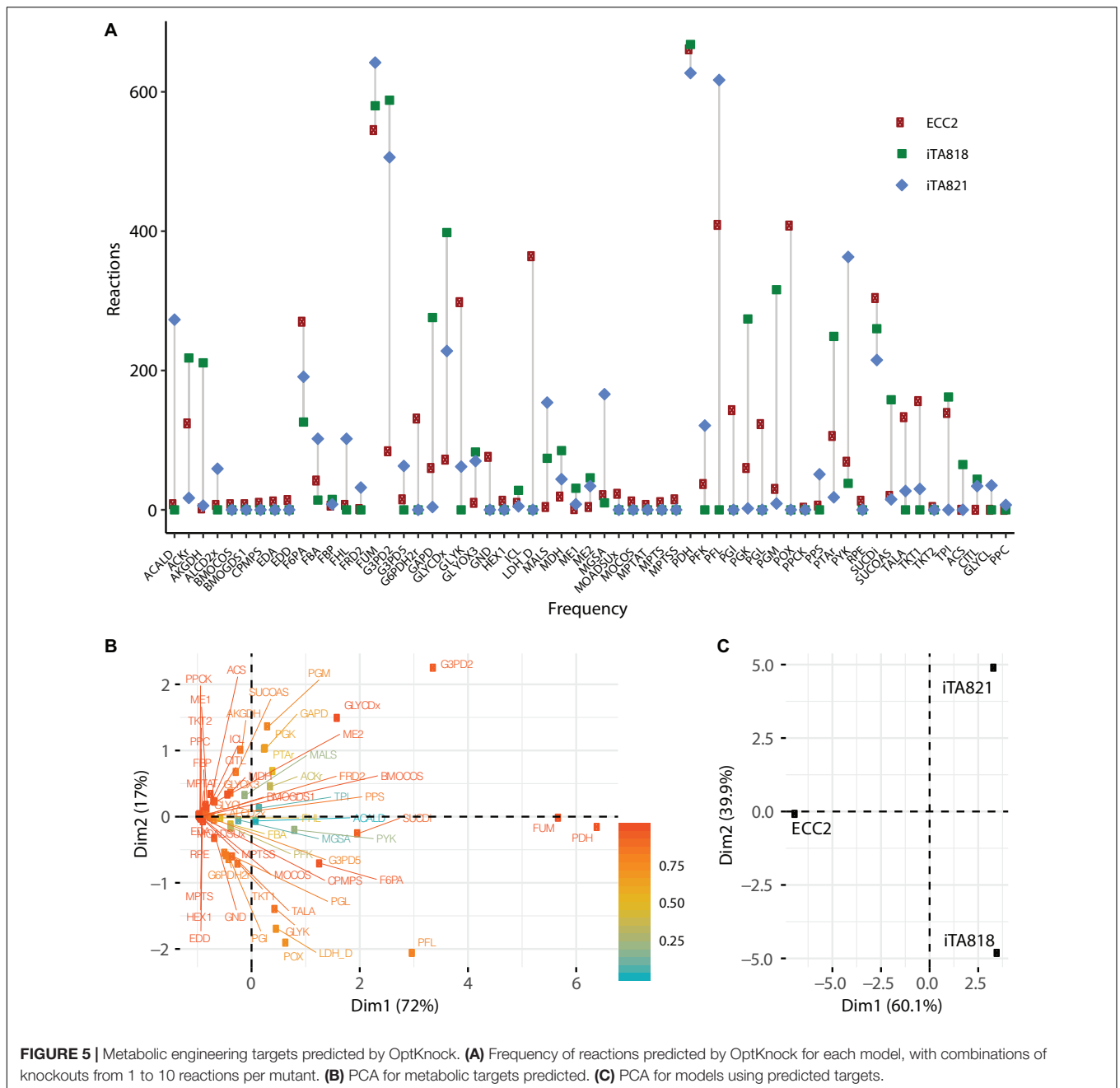
the set of potential reactions to be knocked out. Next, all reactions that had been experimentally found to be essential for growth were removed from consideration (Joyce et al., 2006). Also, the reactions that were found to be computationally essential were not considered, as well as non-gene-associated 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 the ECC2, iTA818, and iTA821 models, respectively (**Supplementary File 2**). **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 frequency. Reactions acetate kinase (ACKr), fructose 6-phosphate aldolase (F6PA), fumarase (FUM), pyruvate dehydrogenase (PDH), pyruvate formate lyase (PFL), phosphotransacetylase (PTAr), succinate dehydrogenase (SUCDi), triosephosphate isomerase (TPI), glycerol-3-phosphate dehydrogenase-NADP (G3PD2), and glycerol dehydrogenase (GLYCDx) were frequently predicted for the 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 the models (**Figure 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

FIGURE 4 | Models comparison after transcriptomic integration using gene inactivity moderated by metabolism and expression (GIMME) under aerobic conditions. **(A)** Venn diagram or reactions included in the model. **(B)** Flux Balance Analysis under aerobic conditions.

FIGURE 5 | Metabolic engineering targets predicted by OptKnock. **(A)** Frequency of reactions predicted by OptKnock for each model, with combinations of knockouts from 1 to 10 reactions per mutant. **(B)** PCA for metabolic targets predicted. **(C)** PCA for models using predicted targets.
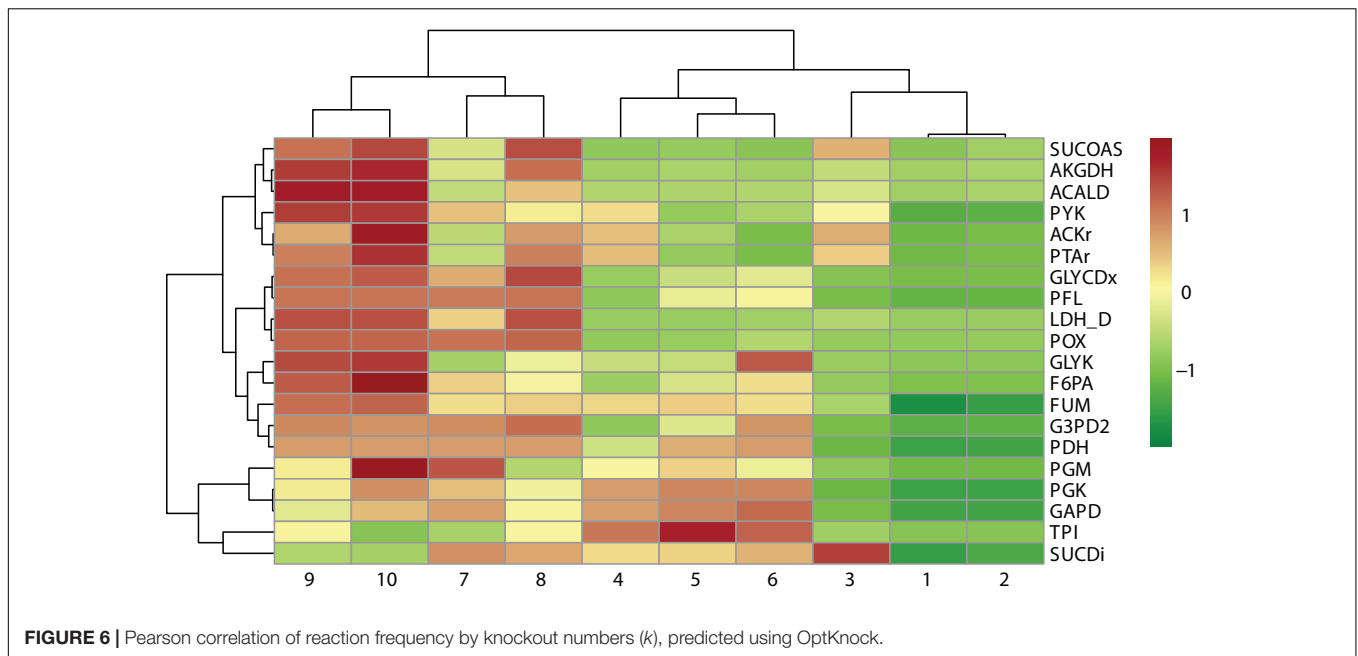
are closely related to succinic acid overproduction from glycerol. Regions of high variability are clustered along with the first principal component, presenting a value of zero for the first principal component. This indicates that the factors that make up the first principal component are critical for high titers. The contributions of different models to the first two principal components of the PCA are shown in **Figure 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 is highly

related to PDH and FUM deletion (**Figure 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 (**Figure 1**), a competitive by-product on succinate production (Blankschien et al., 2010). 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 (Kurzrock and Weuster-Botz, 2010; López-Garzón and Straathof, 2014).

Since metabolic manipulation of cells results in a stressful process, the negative impact of deletions on the maximum growth rate can be observed. To determine the effects of reaction

**FIGURE 6 |** Pearson correlation of reaction frequency by knockout numbers (k), predicted using OptKnock.

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 three to four deletions. The highest succinate production (∼8.5 mmol/g DW h^{-1}) 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 with 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 six 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 six 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, **Figure 7** shows amplifications in the Euclidian distance between the WT and the mutants when succinate production and knockout numbers increase and growth rate decreases.

Identification of Critical Metabolic Targets and Potential Mutants

OptKnock results are a large list of knockout combinations where maximum product synthesis occurs at a maximum growth rate reachable (Burgard et al., 2003). However, it has been observed

that the optimal solution of the target given by OptKnock is not necessarily growth-coupled, and some mutants predicted do not increase the product target. Consequently, selecting 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, these mutants can ensure high productivity over time and initially solve this situation (Shabestary and Hudson, 2016). To identify growth-coupled production solutions, a COBRA Toolbox function was used to verify the minimum and maximum production rates given a set of reactions to be knocked out. As a result, the same minimum and maximum flux for the desired product should be obtained when the maximum growth rate is achieved. One thousand seven hundred ninety-nine (1,799) 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 (**Figure 7**), where 279 mutants were predicted to have a difference between the maximum production rate predicted by the function and FBA < 2, resulting in 2,078 *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 Euclidean distance obtained by MOMA. The dataset 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

FIGURE 7 | Relationship between the number of knockouts, succinic acid production, growth rate, and flux difference.
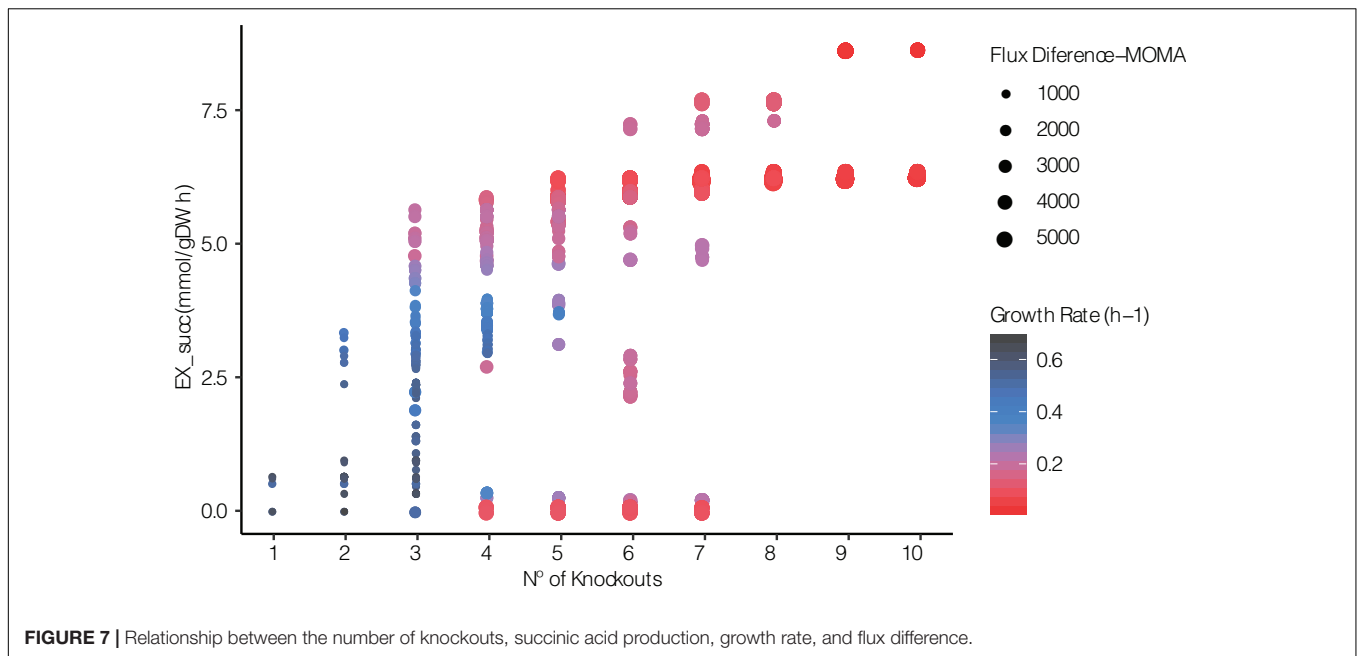
acid production ratio was obtained. A measure of the importance of the contribution of each feature to the random forest model is shown in **Figure 8** indicated by *IncNodePurity*. This model exhibited a mean square error (MSE) value of 0.293 when using the reaction flux of EX_succ_e flux obtained by FBA as a variable response. For 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 9,175.158, 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 the use of machine learning models to predict the largest number of mutants than those obtained by OptKnock in terms of growth rate and succinic acid production since OptKnock is more 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 (**Figure 1**) in glycerol utilization (Blankschien et al., 2010), 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 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 on growth rate (**Figure 8B**).
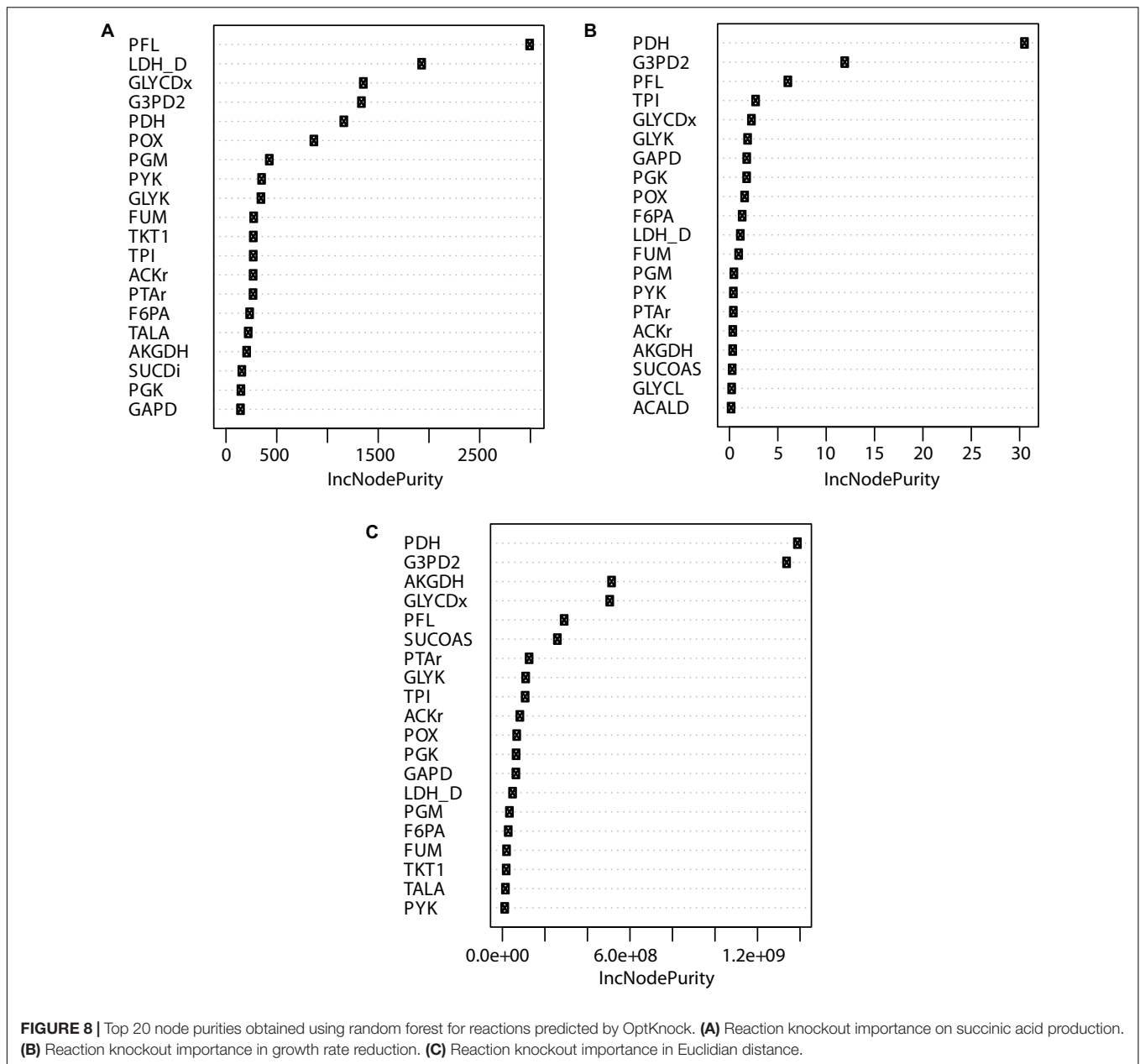
The pyruvate dehydrogenase complex is a critical connection point between glycolysis and the TCA cycle, both of which function during aerobic respiration through catalyzing the conversion of pyruvate to acetyl coenzyme A (acetyl-CoA) (Schutte et al., 2015). PDH deactivation results in PFL carrying the flux from pyruvate to acetyl-CoA (Khodayari et al., 2015). Simple reaction knockouts show that PDH deletion results in a

growth rate reduction of ~5%. Additionally, five reactions (FUM, GAPD, PGK, PGM, and TPI) were predicted to have the most significant reduction (8–10%) in growth rate during glycerol utilization. Of those reactions, only FUM has a significant positive effect over succinate production when this deletion was carried out alone. However, in mutants in which both FUM and PDH were predicted (59.45%), TPI appeared in around 12.60% (**Figure 5**). Then, the deletion of genes associated with TPI in addition to FUM and PDH reactions could negatively affect growth rate.

DISCUSSION

Glycerol metabolism in *E. coli* has been described in the literature (Murarka et al., 2008; Booth, 2014). However, cell changes are carried out as a response to stressful situations. In this study, two conditions were tested for transcription response in *E. coli* to further integrate to metabolic network modeling. Gene expression-wide analyses reveal how cells have the ability to avoid glycerol toxicity, increasing consumption. The most striking response to glycerol consumption and the possible mechanism to optimize succinic acid production from glycerol were revealed by the combination of the 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) (**Figure 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 yield DHAP as well. In this study, overexpression of *glpK* was

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.

observed in both conditions, with a difference of around 20%. This result is not surprising since the GlpK-mediated reaction is a rate-limiting step in glycerol utilization (Herring et al., 2006). However, it has been observed that under the optimized culture conditions, the glycerol utilization rate is higher than that in the evolved conditions, suggesting that other mechanisms should exit 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 between the biosynthesis and catabolism of lipids; however, accumulation of Gly-3-P can become toxic. Thus, it is carefully regulated (Booth, 2014). The export of Gly-3-P could be mediated by *phoE* and *ompF* membrane porins; however, downregulation of *phoE* (−8.67 and −9.04 log2 FC for the optimized culture and the evolved strain, respectively) and

upregulation 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 upregulation 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 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 (Wong and Kwan, 1992; Lemieux et al., 2005). This protein exchanges Gly-3-P for phosphate, avoiding the toxicity of both Gly-3-P and the inorganic phosphates (Booth, 2014). As a result and considering that phosphate is necessary to increase glycerol utilization, autoregulation of the PhoB/PhoR two-component

regulatory system needs to be down-expressed. Downregulation of PhoB/PhoR was observed in this study, which could explain the achievement of optimal density (Gao and Stock, 2013), as well as contribute to the regulation of glycerol phosphate metabolism (Baek and Lee, 2007).

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 (Blankschien et al., 2010; Booth, 2014). In the presence of oxygen or nitrates, GlpD transfers electrons to the respective terminal oxidized. In contrast, under anaerobic conditions, the GlpABC system transfers the electrons to fumarate or nitrates (Unden and Bongaerts, 1997). GlpD upregulation was expected since culture conditions were under aerobic conditions, but a higher expression of the *glpABC* system was surprising. Overexpression of *glpABC* under aerobic conditions could be elucidated because of the activation of fumarate reductase enzymes (**Table 1**) in the evolved strain as a result 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 (Murarka et al., 2008). Moreover, its activity could be supported by *sdhABCD* since they are structurally and functionally homologous (Guest, 1981). Therefore, we hypothesized that *frdABCD* upregulation could be the reason why enhancement in glycerol utilization was not observed in the evolved strain, even when an optimized culture medium was employed.

Insights on the molecular adaptive responses of *E. coli* to glycerol consumption revealed by the transcriptional datasets identified a marked *hdeAB* upregulation only in the evolved strain. This is attractive since HdeAB are periplasmic proteins that play a role in optimal protection at low pH (Masuda and Church, 2003; Kern R. et al., 2007). Therefore, differences in *hdeAB* upregulation in the evolved strain and the optimized culture medium probably occur because acetate is the main product in glycerol utilization, and under ALE conditions, 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.

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 (Velur Selvamani et al., 2014). Then, the 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. The 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 was 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. These results and those obtained in the transcriptional responses suggest that deletion of the *pta* needs to be, almost as mandatory, carried out since acetate production becomes a competitive pathway in glycerol metabolism for succinic acid production (Zhang et al., 2010).

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) (Schutte et al., 2015). PDH deactivation results in PFL carrying the flux from pyruvate to acetyl-CoA (Khodayari et al., 2015). Simple reaction knockouts show that PDH deletion results in a growth rate reduction of ~5%. Additionally, five reactions (FUM, GAPD, PGK, PGM, and TPI) were predicted to have the most significant reduction (8–10%) in growth rate during glycerol utilization. Of those reactions, only FUM has a significant positive effect over succinate production when this deletion was carried out alone. These results indicate that those mutants predicted by OptKnock, where FUM and PDH are predicted, need to be tested in the lab because it has been observed that a low growth rate could negatively affect the profitability of industrial bio-based production products (Chen et al., 2013a; Tafur Rangel et al., 2018). However, in mutants in which both FUM and PDH were predicted (59.45%), TPI appeared in around 12.60% (**Figure 5**). Then, the deletion of genes associated with TPI in addition to FUM and PDH reactions could negatively affect the growth rate. This is because in the absence of TpiA, DHAP is converted to methylglyoxal, which, even at submillimolar concentrations, is a toxic compound (Booth, 2014). DHAP is the result of the alternative pathway on glycerol metabolization consisting 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 (Velur Selvamani et al., 2014). 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 six to seven 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 six to eight reactions are knocked out 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 for the genetic changes carried out by the fewest changes in gene regulation until it achieves an optimal state that could be predicted using FBA (Senger et al., 2015). Then, FBA in engineered strains predicts a long-term evolved state. Thus, an alternative to evaluate unevolved mutants is the MOMA method (Segre et al., 2002). 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 require to be evolutionarily engineered. ALE studies have shown to provide the cell with the ability to grow under selection pressure to go up from a suboptimal state to optimal growth rate predicted using *in silico* models (Ibarra et al., 2002). 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 (Shabestary and Hudson, 2016) by using ALE rounds after metabolically engineering cells (Graf et al., 2019).

CONCLUSION

By adopting tools from various disciplines, computational methods for systems metabolic engineering have been developed to understand cell behavior and how level systems (RNA, proteins, and metabolites, among others) can interact inside the cell for industrial purposes. In the same way, *E. coli* has been extensively studied to become 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 can 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 reduces time in rational strain design and guides in the selection of metabolic engineering targets based on cell behavior under experimental conditions. Simultaneously,

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.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. The RNA datasets presented in this study can be found in the NCBI's Gene Expression Omnibus database and are accessible through GEO Series accession number GSE140847. Further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

AT: conceptualization, methodology, software, validation, investigation, visualization, formal analysis, and writing—original draft preparation. WR, DM, and CO: investigation. RC: methodology, software, and writing—review and editing. JG: supervision and writing—review and editing. AG: conceptualization, supervision, and writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.633073/full#supplementary-material

REFERENCES

Ataman, M., Hernandez Gardiol, D. F., Fengos, G., and Hatzimanikatis, V. (2017). redGEM: Systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models. Senger RS, editor. *PLoS Comput. Biol.* 13:e1005444. doi: 10.1371/journal.pcbi.1005444

Baek, J. H., and Lee, S. Y. (2007). Transcriptome analysis of phosphate starvation response in *Escherichia coli. J. Microbiol. Biotechnol.* 17, 244–252.

Bao, H., Liu, R., Liang, L., Jiang, Y., Jiang, M., Ma, J., et al. (2014). Succinic acid production from hemicellulose hydrolysate by an *Escherichia coli* mutant obtained by atmospheric and room temperature plasma and adaptive evolution. *Enzyme Microb. Technol.* 66, 10–15. doi: 10.1016/j.enzmictec.2014.04.017

Becker, S. A., and Palsson, B. O. (2008). Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.* 4:e1000082. doi: 10.1371/journal.pcbi.1000082

Blankschien, M. D., Clomburg, J. M., and Gonzalez, R. (2010). Metabolic engineering of *Escherichia coli* for the production of succinate from glycerol. *Metab. Eng.* 12, 409–419. doi: 10.1016/j.ymben.2010.06.002

Blazier, A. S., and Papin, J. A. (2012). Integration of expression data in genome-scale metabolic network reconstructions. *Front. Physiol.* 3:299. doi: 10.3389/fphys.2012.00299

Booth, I. R. (2014). Glycerol and methylglyoxal metabolism. *EcoSal. Plus* 1, 1–8.

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

Buschke, N., Schäfer, R., Becker, J., and Wittmann, C. (2013). Metabolic engineering of industrial platform microorganisms for biorefinery applications - Optimization of substrate spectrum and process robustness by rational and evolutive strategies. *Bioresour. Technol.* 135, 544–554. doi: 10.1016/j.biortech.2012.11.047

Carlson, R., Fell, D., and Srienc, F. (2002). Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnol. Bioeng.* 79, 121–134. doi: 10.1002/bit.10305

Cecchini, G., Schröder, I., Gunsalus, R. P., and Maklashina, E. (2002). Succinate dehydrogenase and fumarate reductase from *Escherichia coli*. *Biochim. Biophys. Acta - Bioenerg.* 1553, 140–157. doi: 10.1016/s0005-2728(01)00238-9

Chagneau, C., Heyde, M., Alonso, S., Portalier, R., and Laloi, P. (2001). 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.* 183, 5675–5683. doi: 10.1128/jb.183.19.5675-5683.2001

Chen, X., Xu, G., Xu, N., Zou, W., Zhu, P., Liu, L., et al. (2013b). Metabolic engineering of Torulopsis glabrata for malate production. *Metab. Eng.* 19, 10–16. doi: 10.1016/j.ymben.2013.05.002

Chen, X., Zhou, L., Tian, K., Kumar, A., Singh, S., Prior, B. A., et al. (2013a). Metabolic engineering of *Escherichia coli*: a sustainable industrial platform for bio-based chemical production. *Biotechnol. Adv.* 31, 1200–1223. doi: 10.1016/j.biotechadv.2013.02.009

Chen, Y. P., Lin, H. H., Yang, C. D., Huang, S. H., and Tseng, C. P. (2012). Regulatory role of cAMP receptor protein over *Escherichia coli* fumarase genes. *J. Microbiol.* 50, 426–433. doi: 10.1007/s12275-012-1542-6

Conrad, T. M., Lewis, N. E., and Palsson, B. O. (2011). Microbial laboratory evolution in the era of genome-scale science. *Mol. Syst. Biol.* 7:509. doi: 10.1038/msb.2011.42

Edgar, R. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30, 207–210.

Feist, A. M., Zielinski, D. C., Orth, J. D., Schellenberger, J., Herrgard, M. J., and Palsson, B. O. (2010). Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. *Metab. Eng.* 12, 173–186. doi: 10.1016/j.ymben.2009.10.003

Fong, S. S., Joyce, A. R., and Palsson, B. Ø (2005). Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth. *Genome Res.* 15, 1365–1372. doi: 10.1101/gr.3832305

Fong, S. S., and Marciniak, J. Y. (2003). Description and interpretation of adaptive evolution of *Escherichia coli* K-12 MG1655 by using a genome-scale in silico metabolic model. *Society* 185, 6400–6408. doi: 10.1128/jb.185.21.6400-6408.2003

Förster, A. H., and Gescher, J. (2014). Metabolic engineering of *Escherichia coli* for production of mixed-acid fermentation end products. *Front. Bioeng. Biotechnol.* 2:16. doi: 10.3389/fbioe.2014.00016

Furusawa, C., Horinouchi, T., Hirasawa, T., and Shimizu, H. (2013). Systems metabolic engineering: the creation of microbial cell factories by rational metabolic design and evolution. *Adv. Biochem. Eng. Biotechnol.* 131, 1–23. doi: 10.1007/10_2012_137

Gao, R., and Stock, A. M. (2013). Evolutionary tuning of protein expression levels of a positively autoregulated two-component system. *PLoS Genet.* 9:e1003927. doi: 10.1371/journal.pgen.1003927

Götz, S., García-Gómez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176

Graf, M., Haas, T., Müller, F., Buchmann, A., Harm-Bekbenbetova, J., Freund, A., et al. (2019). Continuous adaptive evolution of a fast-growing corynebacterium glutamicum strain independent of protocatechuate. *Front. Microbiol.* 10:1648. doi: 10.3389/fmicb.2019.01648

Guest, J. R. (1981). Partial replacement of succinate dehydrogenase function by phage- and plasmid-specified fumarate reductase in *Escherichia coli*. *J. Gen. Microbiol.* 122, 171–179. doi: 10.1099/00221287-122-2-171

Haggart, C. R., Bartell, J. A., Saucerman, J. J., and Papin, J. A. (2011). Whole-genome metabolic network reconstruction and constraint-based modeling. *Methods Enzymol.* 500, 411–433. doi: 10.1016/b978-0-12-385118-5.00021-9

Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., et al. (2019). Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat. Protoc.* 14, 639–702.

Herring, C. D., Raghunathan, A., Honisch, C., Patel, T., Applebee, M. K., Joyce, A. R., et al. (2006). Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat. Genet.* 38, 1406–1412. doi: 10.1038/ng1906

Ibarra, R. U., Edwards, J. S., and Palsson, B. O. (2002). *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature* 420, 186–189. doi: 10.1038/nature01149

Jones, H. M., and Gunsalus, R. P. (1987). Regulation of *Escherichia coli* fumarate reductase (frdABCD) operon expression by respiratory electron acceptors and the fnr gene product. *J. Bacteriol.* 169, 3340–3349. doi: 10.1128/jb.169.7.3340-3349.1987

Joyce, A. R., Reed, J. L., White, A., Edwards, R., Osterman, A., Baba, T., et al. (2006). Experimental and computational assessment of conditionally essential genes in *Escherichia coli*. *J. Bacteriol.* 188, 8259–8271. doi: 10.1128/jb.00740-06

Kern, A., Tilley, E., Hunter, I. S., Legiša, M., and Glieder, A. (2007). Engineering primary metabolic pathways of industrial micro-organisms. *J. Biotechnol.* 129, 6–29. doi: 10.1016/j.jbiotec.2006.11.021

Kern, R., Malki, A., Abdallah, J., Tagourti, J., and Richarme, G. (2007). *Escherichia coli* HdeB is an acid stress chaperone. *J. Bacteriol.* 189, 603–610. doi: 10.1128/jb.01522-06

Khodayari, A., Chowdhury, A., and Maranas, C. D. (2015). Succinate overproduction: a case study of computational strain design using a comprehensive *Escherichia coli* kinetic model. *Front. Bioeng. Biotechnol.* 2:76. doi: 10.3389/fbioe.2014.00076

Kim, J. (2012). *Development and Applications of Integrated Metabolic and Transcriptional Regulatory Network Models*. Madison, WI: University of Wisconsin–Madison.

Kim, Y. M., Cho, H. S., Jung, G. Y., and Park, J. M. (2011). Engineering the pentose phosphate pathway to improve hydrogen yield in recombinant *Escherichia coli*. *Biotechnol. Bioeng.* 108, 2941–2946. doi: 10.1002/bit.23259

King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., et al. (2016). BiGG Models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* 44(D1), D515–D522. doi: 10.1093/nar/gkv1049

Kurzrock, T., and Weuster-Botz, D. (2010). Recovery of succinic acid from fermentation broth. *Biotechnol. Lett.* 32, 331–339. doi: 10.1007/s10529-009-0163-6

Lee, D. H., and Palsson, B. O. (2010). Adaptive evolution of escherichia coli K-12 MG1655 during growth on a nonnative carbon source, L-l,2-propanediol. *Appl. Environ. Microbiol.* 76, 4158–4168. doi: 10.1128/aem.00373-10

Lee, J. W., Na, D., Park, J. M., Lee, J., Choi, S., and Lee, S. Y. (2012). Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat. Chem. Biol.* 8, 536–546. doi: 10.1038/nchembio.970

Lemieux, M. J., Huang, Y., and Wang, D. N. (2005). Crystal structure and mechanism of GlpT, the glycerol-3-phosphate transporter from E. coli. *J. Electron. Microsc. (Tokyo)*. 54(Suppl. 1), 43–46.

Li, J., Poursat, M. A., Drubay, D., Motz, A., Saci, Z., Morillon, A., et al. (2015). A dual model for prioritizing cancer mutations in the non-coding genome based on germline and somatic events. *PLoS Comput Biol*. 11:e1004583. doi: 10.1371/journal.pcbi.1004583

Lin, H., Bennett, G. N., and San, K.-Y. (2005). Metabolic engineering of aerobic succinate production systems in *Escherichia coli* to improve process productivity and achieve the maximum theoretical succinate yield. *Metab. Eng.* 7, 116–127. doi: 10.1016/j.ymben.2004.10.003

López-Garzón, C. S., and Straathof, A. J. J. (2014). Recovery of carboxylic acids produced by fermentation. *Biotechnol. Adv.* 32, 873–904. doi: 10.1016/j.biotechadv.2014.04.002

Machado, D., and Herrgård, M. (2014). Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol*. 10:e1003580. doi: 10.1371/journal.pcbi.1003580

Makino, K., Shinagawa, H., Amemura, M., Kawamoto, T., Yamada, M., and Nakata, A. (1989). Signal transduction in the phosphate regulon of *Escherichia coli* involves phosphotransfer between PhoR and PhoB proteins. *J. Mol. Biol.* 210, 551–559. doi: 10.1016/0022-2836(89)90131-9

Masuda, N., and Church, G. M. (2003). Regulatory network of acid resistance genes in *Escherichia coli*. *Mol. Microbiol.* 48, 699–712. doi: 10.1046/j.1365-2958.2003.03477.x

Monk, J. M., Charusanti, P., Aziz, R. K., Lerman, J. A., Premyodhin, N., Orth, J. D., et al. (2013). Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20338–20343. doi: 10.1073/pnas.1307797110

Monk, J. M., Koza, A., Campodonico, M. A., Machado, D., Seoane, J. M., Palsson, B. O., et al. (2016). Multi-omics quantification of species variation of *Escherichia coli* links molecular features with strain phenotypes. *Cell Syst.* 3, 238–251.e12.

Monk, J. M., Lloyd, C. J., Brunk, E., Mih, N., Sastry, A., King, Z., et al. (2017). iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* 35, 904–908. doi: 10.1038/nbt.3956

Murarka, A., Dharmadi, Y., Yazdani, S. S., and Gonzalez, R. (2008). Fermentative utilization of glycerol by *Escherichia coli* and its implications for the production of fuels and chemicals. *Appl. Environ. Microbiol.* 74, 1124–1135. doi: 10.1128/aem.02192-07

Nordlander, B., Krantz, M., and Hohmann, S. (2008). Hog1-mediated metabolic adjustments following hyperosmotic shock in the yeast. *Current* 20, 51–79. doi: 10.1007/4735_2007_0256

O'Brien, E. J., Monk, J. M., and Palsson, B. O. (2015). Using genome-scale models to predict biological capabilities. *Cell* 161, 971–987. doi: 10.1016/j.cell.2015.05.019

Pharkya, P., Burgard, A. P., and Maranas, C. D. (2003). Exploring the overproduction of amino acids using the bilevel optimization framework OptKnock. *Biotechnol. Bioeng.* 84, 887–899. doi: 10.1002/bit.10857

Pharkya, P., Burgard, A. P., and Maranas, C. D. (2004). OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.* 14, 2367–2376. doi: 10.1101/gr.2872004

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. doi: 10.1016/j.ymben.2005.08.003

Portela, C., Villas-Bôas, S., Rocha, I., and Ferreira, E. C. (2013). Genome scale metabolic network reconstruction of the pathogen Enterococcus faecalis. *IFAC Proc. Volumes* 46, 131–136. doi: 10.3182/20131216-3-in-2044.00067

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:e1000744. doi: 10.1371/journal.pcbi.1000744

Rangel, A. E. T., Gómez Ramírez, J. M., and González Barrios, A. F. (2020). From industrial by-products to value-added compounds: the design of efficient microbial cell factories by coupling systems metabolic engineering and bioprocesses. *Biofuels Bioprod. Biorefin.* 14, 1228–1238. doi: 10.1002/bbb.2127

Ruckerbauer, D. E., Jungreuthmayer, C., and Zanghellini, J. (2015). Predicting genetic engineering targets with Elementary Flux Mode Analysis: a review of four current methods. *N. Biotechnol.* 32, 534–546. doi: 10.1016/j.nbt.2015.03.017

Schutte, K. M., Fisher, D. J., Burdick, M. D., Mehrad, B., Mathers, A. J., Mann, B. J., et al. (2015). *Escherichia coli* pyruvate dehydrogenase complex is an important component of CXCL10-mediated antimicrobial activity. *Infect. Immun.* 84, 320–328. doi: 10.1128/iai.00552-15

Segre, D., Vitkup, D., and Church, G. M. (2002). Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15112–15117. doi: 10.1073/pnas.232349399

Senger, R., Yen, J., Tanniche, I., Fisher, A., Gillaspy, G., and Bevan, D. (2015). Designing metabolic engineering strategies with genome-scale metabolic flux modeling. *Adv. Genomics Genet.* 5:93. doi: 10.2147/agg.s58494

Shabestary, K., and Hudson, E. P. (2016). Computational metabolic engineering strategies for growth-coupled biofuel production by Synechocystis. *Metab. Eng. Commun.* 3, 216–226. doi: 10.1016/j.meteno.2016.07.003

Shams Yazdani, S., and Gonzalez, R. (2008). Engineering *Escherichia coli* for the efficient conversion of glycerol to ethanol and co-products. *Metab. Eng.* 10, 340–351. doi: 10.1016/j.ymben.2008.08.005

Tafur Rangel, A. E., Camelo Valera, L. C., Gómez Ramírez, J. M., and González Barrios, A. F. (2018). 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.* 93, 2011–2020. doi: 10.1002/jctb.5432

Unden, G., and Bongaerts, J. (1997). Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. *Biochim. Biophys. Acta - Bioenerg.* 1320, 217–234. doi: 10.1016/s0005-2728(97)00034-0

Varet, H., Brillet-Guéguen, L., Coppée, J. Y., and Dillies, M. A. (2016). SARTools: a DESeq2- and edgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. Mills K, editor. *PLoS One* 11:e0157022. doi: 10.1371/journal.pone.0157022

Velur Selvamani, R. S., Telaar, M., Friehs, K., and Flaschel, E. (2014). Antibiotic-free segregational plasmid stabilization in *Escherichia coli* owing to the knockout of triosephosphate isomerase (tpiA). *Microb. Cell Fact.* 13:58. doi: 10.1186/1475-2859-13-58

Vlysidis, A., Binns, M., Webb, C., and Theodoropoulos, C. (2011). A techno-economic analysis of biodiesel biorefineries: assessment of integrated designs for the co-production of fuels and chemicals. *Energy* 36, 4671–4683. doi: 10.1016/j.energy.2011.04.046

Wang, Y., Manow, R., Finan, C., Wang, J., Garza, E., and Zhou, S. (2011). Adaptive evolution of nontransgenic *Escherichia coli* KC01 for improved ethanol tolerance and homoethanol fermentation from xylose. *J. Ind. Microbiol. Biotechnol.* 38, 1371–1377. doi: 10.1007/s10295-010-0920-5

Wong, K. K., and Kwan, H. S. (1992). Transcription of glpT of *Escherichia coli* K12 is regulated by anaerobiosis and fnr. *FEMS Microbiol. Lett.* 94, 15–18.

Woo, H. M., and Park, J. B. (2014). Recent progress in development of synthetic biology platforms and metabolic engineering of Corynebacterium glutamicum. *J. Biotechnol.* 180, 43–51. doi: 10.1016/j.jbiotec.2014.03.003

Yin, X., Li, J., Shin, H. D., Du, G., Liu, L., and Chen, J. (2015). Metabolic engineering in the biotechnological production of organic acids in the tricarboxylic acid cycle of microorganisms: advances and prospects. *Biotechnol. Adv.* 33, 830–841. doi: 10.1016/j.biotechadv.2015.04.006

Zhang, B., Skory, C. D., and Yang, S. T. (2012). Metabolic engineering of Rhizopus oryzae: effects of overexpressing pyc and pepc genes on fumaric acid biosynthesis from glucose. *Metab. Eng.* 14, 512–520. doi: 10.1016/j.ymben.2012.07.001

Zhang, J., Wu, C., Du, G., and Chen, J. (2012). Enhanced acid tolerance in Lactobacillus casei by adaptive evolution and compared stress response during acid stress. *Biotechnol. Bioprocess Eng.* 17, 283–289.

Zhang, X., Shanmugam, K. T., and Ingram, L. O. (2010). Fermentation of glycerol to succinate by metabolically engineered strains of *escherichia coli*. *Appl. Environ. Microbiol.* 76, 2397–2401.

Zhu, Q., and Jackson, E. N. (2015). Metabolic engineering of Yarrowia lipolytica for industrial applications. *Curr. Opin. Biotechnol.* 36, 65–72. doi: 10.1016/j.copbio.2015.08.010